国家出版基金项目
NATIONAL PUBLICATION FOUNDATION

# THE SAGE ENCYCLOPEDIA OF SOCIAL SCIENCE RESEARCH METHODS

# 社会科学研究方法百科全书

【第一卷】

[美]米歇尔·刘易斯-伯克（Michael S. Lewis-Beck）
[美]艾伦·布里曼（Alan Bryman）
[美]廖福挺（Tim Futing Liao）—主编

沈崇麟　赵　锋　高　勇—主译

沈崇麟　杨　可　郑晓娟—本卷译者

重庆大学出版社

SHEHUI KEXUE YANJIU FANGFA BAIKE QUANSHU

更多服务

ISBN 978-7-5689-0785-9

9 787568 907859 >

总定价：698.00元（共三卷）

国家出版基金项目
NATIONAL PUBLICATION FOUNDATION

# 社会科学研究方法
# 百科全书

【第一卷】

[美]米歇尔·刘易斯-伯克（Michael S. Lewis-Beck）
[美]艾伦·布里曼（Alan Bryman）
[美]廖福挺（Tim Futing Liao） 主编

沈崇麟 赵 锋 高 勇 主译

沈崇麟 杨 可 郑晓娟 本卷译者

译者（按姓氏拼音排序）

高 勇（中国社会科学院） 李 洋（北京市社会科学院） 马 妍（中国社会科学院）
沈崇麟（中国社会科学院） 王 玥（首都师范大学） 杨 可（中国社会科学院）
叶鹏飞（中国劳动关系学院）赵 锋（中国社会科学院） 郑晓娟（北京大学）

词条审定专家（按姓氏拼音排序）

郭金华（北京大学） 郭志刚（北京大学） 何江穗（中国政法大学）
柯惠新（中国传媒大学） 李 原（中国社会科学院） 渠敬东（北京大学）
沈崇麟（中国社会科学院）席仲恩（重庆邮电大学） 夏传玲（中国社会科学院）

重庆大学出版社

版贸核渝字(2012)第094号

**图书在版编目(CIP)数据**

社会科学研究方法百科全书. 第一卷 / (美) 米歇尔·刘易斯-伯克, (美) 艾伦·布里曼, (美) 廖福挺主编; 沈崇麟,赵锋,高勇主译. -- 重庆 : 重庆大学出版社, 2017.8
(万卷方法)
书名原文:The SAGE Encyclopedia of Social Science Research Methods
ISBN 978-7-5689-0785-9

Ⅰ. ①社… Ⅱ. ①米… ②艾… ③廖… ④沈… ⑤赵… ⑥高… Ⅲ.①社会科学—研究方法 Ⅳ. ①C3

中国版本图书馆 CIP 数据核字(2017)第199789号

社会科学研究方法百科全书
(第一卷)
[美]米歇尔·刘易斯-伯克 [美]艾伦·布里曼 [美]廖福挺 主编
沈崇麟 赵 锋 高 勇 主译
策划编辑:雷少波 柏子康 林佳木
责任编辑:雷少波 柏子康 屈腾龙 张慧梓 姜 凤 版式设计:雷少波
责任校对:张红梅 责任印制:张 策
*
重庆大学出版社出版发行
出版人:易树平
社址:重庆市沙坪坝区大学城西路21号
邮编:401331
电话:(023)88617190 88617185(中小学)
传真:(023)88617186 88617166
网址:http://www.cqup.com.cn
邮箱:fxk@cqup.com.cn (营销中心)
全国新华书店经销
重庆升光电力印务有限公司印刷
*
开本:787mm×1092mm 1/16 印张:35 字数:806千
2017年8月第1版 2017年8月第1次印刷
ISBN 978-7-5689-0785-9 总定价:698.00元(共三卷)

---

# 撰稿人

**Abdi, Hervé**
University of Texas, Dallas
**Adair, John G.**
University of Manitoba, Canada
**Aditya, Ram N.**
Florida International University, Miami
**Adler, Peter**
University of Denver, Colorado
**Adler, Patricia A.**
University of Colorado, Boulder
**Aiken, Leona S.**
Arizona State University, Tempe
**Albritton, Robert B.**
University of Mississippi
**Aldridge, Alan**
University of Nottingham, United Kingdom
**Allison, Paul D.**
University of Pennsylvania, Philadelphia
**Altheide, David L.**
Arizona State University, Tempe
**Altman, Micah**
Harvard University, Massachusetts
**Alvesson, Mats**
University of Lund, Sweden
**Amoureux, Jacque L.**
University of Iowa, Iowa City
**Andersen, Robert**
McMaster University, Canada
**Anderson, Carolyn J.**
University of Illinois, Urbana-Champaign
**Angle, John**
US Department of Agriculture, Washington, DC
**Angrosino, Michael V.**
University of South Florida, Tampa
**Anselin, Luc**
University of Illinois, Urbana-Champaign
**Aronow, Edward**
Montclair State University, New Jersey
**Ashmore, Malcolm**
Loughborough University, United Kingdom
**Atkinson, Robert**
University of Southern Maine, Gorham
**Atkinson, Rowland**
University of Glasgow, United Kingdom
**Bacher, Johann**
Universität Erlangen-Nürnberg, Germany
**Baird, Chardie L.**
Florida State University, Tallahassee
**Bakeman, Roger**
Georgia State University, Atlanta
**Bakker, Ryan**
University of Florida, Gainesville
**Baltagi, Badi H.**
Texas A&M University
**Barr, Robert**
University of Manchester, United Kingdom
**Bartunek, Jean M.**
Boston College, Massachusetts
**Batson, C. Daniel**
University of Kansas, Lawrence
**Beck, Nathaniel**
University of California, San Diego; New York University
**Becker, Saul**
Loughborough University, United Kingdom
**Becker, Mark P.**
University of Minnesota, Minneapolis
**Benton, Ted**
University of Essex, United Kingdom
**Berk, Richard A.**
University of California, Los Angeles
**Billig, Michael**
Loughborough University, United Kingdom
**Bishop, George F.**
University of Cincinnati, Ohio
**Black, Thomas R.**
University of Surrey, United Kingdom
**Blaikie, Norman**
Formerly at the University of Science, Malaysia and the RMIT University, Australia
**Blascovich, Jim**
University of California, Santa Barbara
**Boehmke, Frederick J.**
University of Iowa, Iowa City

**Boklund-Lagopoulos, Karin**
Aristotle University of Thessaloniki, Greece
**Bollen, Kenneth A.**
University of North Carolina, Chapel Hill
**Bornat, Joanna**
The Open University, United Kingdom
**Bornstein, Robert F.**
Gettysburg College, Pennsylvania
**Bottorff, Joan L.**
University of British Columbia, Canada
**Bourque, Linda B.**
University of California, Los Angeles
**Box-Steffensmeier, Janet M.**
Ohio State University, Columbus
**Brady, Ivan**
State University of New York, Oswego
**Brennan, Robert L.**
University of Iowa, Iowa City
**Brewer, Marilynn B.**
Ohio State University, Columbus
**Brick, J. Michael**
Westat, Rockville, Maryland
**Brookfield, Stephen D.**
University of St. Thomas, Minneapolis-St. Paul
**Brown, Courtney**
Emory University, Atlanta
**Brown, Steven R.**
Kent State University
**Bryman, Alan**
Loughborough University, United Kingdom
**Buck, Nicholas H.**
University of Essex, United Kingdom
**Burkhart, Ross E.**
Boise State University, Idaho
**Burr, Vivien**
University of Huddersfield, United Kingdom
**Byrne, David**
University of Durham, United Kingdom
**Cameron, Amna**
North Carolina State University, Raleigh
**Carless, Sally A.**
Monash University, Australia
**Carmines, Edward G.**
Indiana University, Bloomington
**Chapin, Tim**
Florida State University, Tallahassee
**Charlton, Tony**
University of Gloucestershire, United Kingdom
**Charlton, John R. H.**
Office for National Statistics, United Kingdom
**Charmaz, Kathy**
Sonoma State University, California
**Chartrand, Tanya L.**
Duke University, North Carolina
**Chell, Elizabeth**
University of Southampton, United Kingdom
**Chen, Peter Y.**
Colorado State University, Fort Collins
**Christ, Sharon L.**
University of North Carolina, Chapel Hill
**Ciambrone, Desirée**
Rhode Island College, Providence
**Civettini, Andrew J.**
University of Iowa, Iowa City
**Clark, M. H.**
The University of Memphis, Tennessee
**Clarke, Harold D.**
University of Texas, Dallas
**Clegg, Chris**
University of Sheffield, United Kingdom
**Clinton, Joshua D.**
Princeton University, New Jersey
**Coffey, Amanda**
Cardiff University, United Kingdom
**Cole, Michael**
University of California, San Diego
**Cooper, Harris M.**
Duke University, North Carolina
**Copp, Martha A.**
East Tennessee State University, Johnson City
**Corbin, Juliet M.**
The University of Alberta, Canada
**Corter, James E.**
Columbia University, New York
**Corti, Louise**
University of Essex, United Kingdom
**Cortina, Jose M.**
George Mason University, Virginia
**Couper, Mick P.**
University of Michigan, Ann Arbor
**Coxon, Anthony P. M.**
University of Edinburgh, United Kingdom
**Craib, Ian**
Late of University of Essex, United Kingdom
**Cramer, Duncan**
Loughborough University, United Kingdom
**Crano, William D.**
Claremont Graduate University, California
**Croll, Paul**
The University of Reading, United Kingdom
**Croon, Marcel A.**
Tilburg University, The Netherlands
**Currivan, Douglas B.**
University of Massachusetts, Boston
**Dale, Angela**
University of Manchester, United Kingdom
**De Boef, Suzanna**
Pennsylvania State University, University Park

**de Vaus, David A.**
La Trobe University, Australia
**Denzin, Norman K.**
University of Illinois, Urbana-Champaign
**Dixon-Woods, Mary**
University of Leicester, United Kingdom
**Doreian, Patrick**
London School of Economics, United Kingdom
**Dovidio, John F.**
Colgate University, New York
**Dressler, William W.**
The University of Alabama, Tuscaloosa
**Drew, Paul**
University of York, United Kingdom
**Easterby-Smith, Mark**
Lancaster University, United Kingdom
**Edwards, Derek**
Loughborough University, United Kingdom
**Edwards, W. Sherman**
Westat, Rockville, Maryland
**Eisenhardt, Kathleen M.**
Stanford University, California
**Eliason, Scott R.**
University of Minnesota, Minneapolis
**Ellis, Charles H.**
Ohio State University, Columbus
**Essed, Philomena**
University of Amsterdam, The Netherlands
**Fairclough, Norman**
Lancaster University, United Kingdom
**Fetterman, David M.**
Stanford University, California
**Filmer, Paul**
Goldsmiths' College, United Kingdom
**Finkel, Steven E.**
University of Virginia, Charlottesville
**Flint, John**
University of Glasgow, United Kingdom
**Foley, Hugh J.**
Skidmore College, New York
**Fontana, Andrea**
University of Nevada, Las Vegas
**Fowler, Jr., Floyd J.**
University of Massachusetts, Boston
**Fox, James Alan**
Northeastern University, Massachusetts
**Fox, John**
McMaster University, Canada
**Franzosi, Roberto**
University of Reading, United Kingdom
**Frederick, Richard I.**
US Medical Center for Federal Prisoners, Missouri
**Freedman, David A.**
University of California, Berkeley
**Frey, James H.**
University of Nevada, Las Vegas
**Friendly, Michael**
York University, Canada
**Fuller, Duncan**
University of Northumbria at Newcastle, United Kingdom
**Furbee, N. Louanna**
University of Missouri, Columbia
**Gallagher, Patricia M.**
University of Massachusetts, Boston
**Gallmeier, Charles P.**
Indiana University Northwest, Gary
**Garson, G. David**
North Carolina State University, Raleigh
**Gephart, Jr., Robert P.**
University of Alberta, Canada
**Gergen, Kenneth J.**
Swarthmore College, Pennsylvania
**Gerhardt, Uta**
Universität Heidelberg, Germany
**Gershuny, Jonathan**
University of Essex, United Kingdom
**Gibbons, Jean D.**
University of Alabama
**Gibbs, Graham R.**
University of Huddersfield, United Kingdom
**Gibson, Nancy**
University of Alberta, Canada
**Gilbert, Nigel**
University of Surrey, United Kingdom
**Gill, Jeff**
University of Florida, Gainesville
**Glasgow, Garrett**
University of California, Santa Barbara
**Glenn, Norval D.**
University of Texas, Austin
**Goldstone, Jack**
University of California, Davis
**Green, Donald P.**
Yale University, Connecticut
**Greenacre, Michael**
Universitat Pomepu Fabra, Spain
**Greenland, Sander**
University of California, Los Angeles
**Greenwood, Davydd J.**
Cornell University, New York
**Griffith, James W.**
Northwestern University, Illinois
**Gross, Alan L.**
City University of New York
**Gschwend, Thomas**
University of Mannheim, Germany
**Guba, Egon G.**
Indiana University

**Gubrium, Jaber F.**
University of Missouri, Columbia
**Gujarati, Damodar N.**
United States Military Academy, West Point
**Guo, Guang**
University of North Carolina, Chapel Hill
**Guy, Will**
University of Bristol, United Kingdom
**Hagenaars, Jacques A.**
Tilburg University, The Netherlands
**Hagle, Timothy M.**
University of Iowa, Iowa City
**Hammersley, Martyn**
The Open University, United Kingdom
**Han, Shin-Kap**
University of Illinois, Urbana-Champaign
**Hancock, Gregory R.**
University of Maryland, College Park
**Handcock, Mark S.**
University of Washington, Seattle
**Hansen, Peter Reinhard**
Brown University, Rhode Island
**Hardy, Melissa A.**
Pennsylvania State University, State College
**Hargens, Lowell L.**
University of Washington, Seattle
**Harper, Douglas**
Duquesne University, Pennsylvania
**Harré, Rom**
University of Oxford, United Kingdom
**Headland, Thomas N.**
SIL International, Dallas, Texas
**Henry, Gary T.**
Georgia State University, Atlanta
**Herron, Michael C.**
Dartmouth College, New Hampshire
**Hessling, Robert M.**
University of Wisconsin, Milwaukee
**Hine, Christine M.**
University of Surrey United Kingdom
**Hipp, John R.**
University of North Carolina, Chapel Hill
**Hocutt, Max**
The University of Alabama, Tuscaloosa
**Hollway, Wendy**
The Open University, United Kingdom
**Holman, Darryl J.**
University of Washington, Seattle
**Holman, David J.**
University of Sheffield, United Kingdom
**Holstein, James A.**
Marquette University, Wisconsin
**Homan, Roger**
University of Brighton, United Kingdom
**Horrace, William C.**
Syracuse University, New York
**House, Ernest R.**
University of Colorado, Boulder
**Hoyle, Rick H.**
Duke University, North Carolina
**Hughes, Rhidian**
King's College London, United Kingdom
**Hundley, Vanora**
University of Aberdeen, United Kingdom
**Iversen, Gudmund R.**
Swarthmore College, Pennsylvania
**Jaccard, James J.**
State University of New York, Albany
**Jackman, Simon**
Stanford University, California
**Jacoby, William G.**
Michigan State University, East Lansing
**Jasso, Guillermina**
New York University
**Jefferis, Valerie E.**
Ohio State University, Columbus
**Jefferson, Tony**
Keele University, United Kingdom
**Jenkins, Richard**
University of Sheffield, United Kingdom
**Johnson, Jeffrey C.**
East Carolina University, Greenville
**Johnson, Paul E.**
University of Kansas, Lawrence
**Jones, Michael Owen**
University of California, Los Angeles
**Kanuha, Valli Kalei**
University of Hawaii, Honolulu
**Kaplan, David W.**
University of Delaware, Newark
**Karuntzos, Georgia T.**
RTI International, Research Triangle Park, North Carolina
**Katz, Jonathan N.**
California Institute of Technology, Pasadena
**Kedar, Orit**
University of Michigan, Ann Arbor
**Kennedy, Peter E.**
Simon Fraser University, Canada
**Kenny, David A.**
University of Connecticut, Storrs
**Kimball, David C.**
University of Missouri, St. Louis
**King, Jean A.**
University of Minnesota, Minneapolis
**Kirk, Roger E.**
Baylor University, Texas
**Knoke, David H.**
University of Minnesota, Minneapolis

**Kolen, Michael J.**
University of Iowa, Iowa City
**Korczynski, Marek**
Loughborough University, United Kingdom
**Kovtun, Mikhail**
Duke University, North Carolina
**Krauss, Autumn D.**
Colorado State University, Fort Collins
**Kroonenberg, Pieter M.**
Leiden University, The Netherlands
**Krueger, Richard A.**
University of Minnesota, St. Paul
**Kuhn, Randall**
University of Colorado, Boulder
**Kuzel, Anton J.**
Virginia Commonwealth University, Richmond
**Lagopoulos, Alexandros**
Aristotle University of Thessaloniki, Greece
**Langeheine, Rolf**
Formerly at the University of Kiel, Germany
**Laurie, Heather**
University of Essex, United Kingdom
**Lauriola, Marco**
University of Rome, "La Sapienza" Italy
**Lavrakas, Paul J.**
Nielsen Media Research, New York
**Le Voi, Martin**
The Open University, United Kingdom
**Leahey, Erin**
University of Arizona, Tucson
**Ledolter, Johannes**
University of Iowa, Iowa City
**Lee, Raymond M.**
Royal Holloway University of London, United Kingdom
**Lee (Zhen Li), Jane**
University of Florida, Gainesville
**Levin, Irwin P.**
University of Iowa, Iowa City
**Lewis-Beck, Michael S.**
University of Iowa, Iowa City
**Liao, Tim Futing**
University of Essex, United Kingdom; University of Illinois, Urbana-Champaign
**Lincoln, Yvonna S.**
Texas A&M University
**Little, Daniel**
University of Michigan, Dearborn
**Little, Jani S.**
University of Colorado, Boulder
**Liu, Cong**
Illinois State University, Normal
**Lockyer, Sharon**
De Montfort University, United Kingdom
**Long, Karen J.**
University of Michigan, Ann Arbor
**Lowe, Will**
University of Bath, United Kingdom
**Lykken, David T.**
University of Minnesota, Minneapolis
**Lynch, Michael**
Cornell University, New York
**Lynn, Peter**
University of Essex, United Kingdom
**Madsen, Douglas**
University of Iowa, Iowa City
**Magidson, Jay**
Statistical Innovations, Belmont, Massachusetts
**Mahoney, Christine**
Pennsylvania State University, University Park
**Mahoney, James**
Brown University, Rhode Island
**Maines, David R.**
Oakland University, Michigan
**Manton, Kenneth G.**
Duke University, North Carolina
**Marsden, Peter V.**
Harvard University, Massachuesetts
**Marsh, Lawrence C.**
University of Notre Dame, Indiana
**Mason, Jennifer**
University of Leeds, United Kingdom
**Mathiowetz, Nancy**
University of Wisconsin, Milwaukee
**Maxim, Paul S.**
University of Western Ontario, Canada
**May, Tim**
University of Salford, United Kingdom
**Maynard, Mary**
University of York, United Kingdom
**McDowall, David**
State University of New York, Albany
**McElhanon, Kenneth A.**
Graduate Institute of Applied Linguistics, Texas
**McGraw, Kenneth O.**
University of Mississippi
**McGuigan, Jim**
Loughborough University, United Kingdom
**McIver, John P.**
University of Colorado, Boulder
**McKelvie, Stuart J.**
Bishop's University, Canada
**Meadows, Lynn M.**
University of Calgary, Canada
**Menard, Scott**
University of Colorado, Boulder
**Merrett, Mike C.**
University of Essex, United Kingdom

**Mertens, Donna M.**
Gallaudet University, Washington, DC
**Miller, Arthur G.**
Miami University, Ohio
**Millward, Lynne**
University of Surrey, United Kingdom
**Moffett, Kenneth W.**
University of Iowa, Iowa City
**Mooney, Christopher Z.**
University of Illinois, Springfield
**Moreno, Erika**
University of Iowa, Iowa City
**Morris, Martina**
University of Washington, Seattle
**Morse, Janice M.**
University of Alberta, Canada
**Moyé, Lemuel A.**
University of Texas, Houston
**Mueller, Charles W.**
University of Iowa, Iowa City
**Mueller, Daniel J.**
Indiana University, Bloomington
**Mueller, Ralph O.**
The George Washington University, Washington, DC
**Mukhopadhyay, Kausiki**
University of Denver, Colorado
**Mulaik, Stanley A.**
Georgia Institute of Technology, Atlanta
**Nagler, Jonathan**
New York University
**Nash, Chris**
University of Technology, Australia
**Nathan, Laura**
Mills College, California
**Nishisato, Shizuhiko**
University of Toronto, Canada
**Noblit, George W.**
University of North Carolina, Chapel Hill
**Norpoth, Helmut**
State University of New York, Stony Brook
**Noymer, Andrew**
University of California, Berkeley
**Oakley, Ann**
University of London, United Kingdom
**Olsen, Randall J.**
Ohio State University, Columbus
**Olsen, Wendy K.**
University of Manchester, United Kingdom
**Ondercin, Heather L.**
Pennsylvania State University, University Park
**Outhwaite, R. William**
University of Sussex, United Kingdom
**Page, Stewart**
University of Windsor, Canada
**Palmer, Harvey D.**
University of Mississippi
**Paluck, Elizabeth Levy**
Yale University, Connecticut
**Pampel, Fred**
University of Colorado, Boulder
**Parcero, Osiris J.**
Bristol University, United Kingdom
**Park, Sung Ho**
Texas A&M University
**Parry, Ken W.**
Griffith University, Australia
**Pattison, Philippa**
University of Melbourne, Australia
**Patton, Michael Quinn**
The Union Institute & University, Utilization-Focused Evaluation, Minnesota
**Paul, Pallab**
University of Denver, Colorado
**Pavlichev, Alexei**
North Carolina State University, Raleigh
**Paxton, Pamela**
Ohio State University, Columbus
**Peng, Chao-Ying Joanne**
Indiana University, Bloomington
**Phillips, Nelson**
University of Cambridge, United Kingdom
**Phua, VoonChin**
City University of New York
**Pickering, Michael J.**
Loughborough University, United Kingdom
**Pink, Sarah**
Loughborough University, United Kingdom
**Platt, Lucinda**
University of Essex, United Kingdom
**Platt, Jennifer**
University of Sussex, United Kingdom
**Plummer, Ken**
University of Essex, United Kingdom
**Poland, Blake D.**
University of Toronto, Canada
**Pollins, Brian M.**
The Mershon Center, Columbus, Ohio
**Post, Alison E.**
Harvard University, Massachusetts
**Potter, Jonathan A.**
Loughborough University, United Kingdom
**Prosch, Bernhard**
Universität Erlangen-Nürnberg, Germany
**Pye, Paul E.**
University of Teesside, United Kingdom
**Quinn, Kevin M.**
Harvard University, Massachusetts

**Ragin, Charles C.**
University of Arizona, Tucson
**Rässler, Susanne**
Universität Erlangen-Nürenberg, Germany
**Redlawsk, David P.**
University of Iowa, Iowa City
**Reis, Harry T.**
University of Rochester, New York
**Richardson, Laurel**
Ohio State University, Columbus
**Riessman, Catherine K.**
Boston University, Massachusetts
**Ritzer, George**
University of Maryland, College Park
**Robinson, Dawn T.**
University of Iowa, Iowa City
**Rodgers, Joseph Lee**
University of Oklahoma, Norman
**Rodríguez, Germán**
Princeton University, New Jersey
**Roncek, Dennis W.**
University of Nebraska, Omaha
**Rosenthal, Robert**
University of California, Riverside
**Rosnow, Ralph L.**
Temple University, Pennsylvania
**Rothstein, Hannah R.**
Baruch College, New York
**Rubin, Donald B.**
Harvard University, Massachusetts
**Rudas, Tamás**
Eotvos Lorand University, Hungary
**Sandelowski, Margarete**
University of North Carolina, Chapel Hill
**Santos, Filipe M.**
INSEAD, France
**Sapsford, Roger**
University of Teesside, United Kingdom
**Sasaki, Masamichi**
Hyogo Kyoiku University, Japan
**Schafer, William D.**
University of Maryland, College Park
**Schelin, Shannon Howle**
North Carolina State University, Raleigh
**Schenker, Nathaniel**
National Center for Health Statistics, Maryland
**Schmidt, Tara J.**
University of Wisconsin, Milwaukee
**Schrodt, Philip A.**
University of Kansas, Lawrence
**Scott, Steven L.**
University of Southern California, Los Angeles
**Scott, John**
University of Essex, United Kingdom
**Seale, Clive**
Brunel University, United Kingdom
**Shadish, William R.**
University of California, Merced
**Shaffir, William**
McMaster University, Canada
**Shannon, Megan L.**
University of Iowa, Iowa City
**Shipan, Charles R.**
University of Iowa, Iowa City
**Shu, Xiaoling**
University of California, Davis
**Sieber, Joan E.**
California State University, Hayward
**Sijtsma, Klaas**
Tilburg University, The Netherlands
**Silver, N. Clayton**
University of Nevada, Las Vegas
**Slagter, Tracy Hoffmann**
University of Iowa, Iowa City
**Smith, Peter W. F.**
University of Southampton, United Kingdom
**Smith, John K.**
University of Northern Iowa, Cedar Falls
**Smithson, Michael**
Australian National University, Australia
**Snijders, Tom A. B.**
University of Groningen, The Netherlands
**Spector, Paul E.**
University of South Florida, Tampa
**Standing, Lionel G.**
Bishop's University, Canada
**Stark, Philip B.**
University of California, Berkeley
**Stevens, Gillian**
University of Illinois, Urbana-Champaign
**Stillman, Todd**
University of Maryland, College Park
**Stokes, S. Lynne**
Southern Methodist University, Texas
**Stovel, Katherine**
University of Washington, Seattle
**Strohmetz, David B.**
Monmouth University, New Jersey
**Stuart, Elizabeth A.**
Harvard University, Massachusetts
**Sutton, Alex**
University of Leicester, United Kingdom
**Sweeney, Kevin J.**
Ohio State University, Columbus
**Swicegood, C. Gray**
University of Illinois, Urbana-Champaign
**Tacq, Jacques**
Catholic University of Brussels, Belgium

**Tcherni, Maria**
Northeastern University, Massachusetts
**Thomas, Jim**
Northern Illinois University, DeKalb
**Thompson, Bruce**
Texas A&M University
**Thorne, Sally E.**
University of British Columbia, Canada
**Tien, Charles**
City University of New York
**Ting, Kwok-fai**
The Chinese University of Hong Kong
**Toothaker, Larry E.**
University of Oklahoma, Norman
**Tracy, Karen**
University of Colorado, Boulder
**Traugott, Michael W.**
University of Michigan, Ann Arbor
**Traxel, Nicole M.**
University of Wisconsin, Milwaukee
**Tsuji, Ryuhei**
Meiji Gakuin University, Japan
**Tuchman, Gaye**
University of Connecticut, Storrs
**Tuma, Nancy Brandon**
Stanford University, California
**Vallet, Louis-André**
LASMAS-Institut du Longitudinal (CNRS) and Laboratoire de Sociologie Quantitative (CREST), France
**van de Pol, Frank J. R.**
Statistics Netherlands, The Netherlands
**van der Ark, L. Andries**
Tilburg University, The Netherlands
**van Manen, Max**
University of Alberta, Canada
**van Teijlingen, Edwin R.**
University of Aberdeen, United Kingdom
**Vann, Irvin B.**
North Carolina State University, Raleigh
**Vermunt, Jeroen K.**
Tilburg University, The Netherlands
**von Hippel, Paul T.**
Ohio State University, Columbus
**Walker, Robert**
University of Nottingham, United Kingdom
**Walsh, Susan**
University of Sheffield, United Kingdom
**Wand, Jonathan**
Harvard University, Massachusetts
**Wang, Cheng-Lung**
Florida State University, Tallahassee
**Warren, Carol A. B.**
University of Kansas, Lawrence
**Wasserman, Stanley**
University of Illinois, Urbana-Champaign
**Weisberg, Herbert F.**
Ohio State University, Columbus
**Weisfeld, Carol C.**
University of Detroit Mercy, Michigan
**Wengraf, Tom**
Middlesex University, United Kingdom
**Whitten, Guy**
Texas A&M University
**Wilcox, Rand R.**
University of Southern California, Los Angeles
**Williams, Malcolm**
University of Plymouth, United Kingdom
**Wood, B. Dan**
Texas A&M University
**Woods, James**
West Virginia University, Morgantown
**Xie, Yu**
University of Michigan, Ann Arbor
**Zimmerman, Marc**
University of Houston, Texas
**Zorn, Christopher**
National Science Foundation, Arlington, Virginia

# 译者前言

《社会科学研究方法百科全书》(以下简称《百科全书》)是关于社会科学研究方法的全方位、教育式的参考集合。本《百科全书》的内容是全方位的,涵盖了方法学的几乎所有领域:研究设计(research design)、测量(measurement)、抽样调查(survey)、回归(regression)和时间序列(time-series)等统计模型,田野研究(field research)、个案研究(case study)等定性研究(qualitative research)方法,评估(evaluation)、实验设计(experimental design)和模拟(simulation)等研究方式,还包括了像女性主义方法论(feminism methodology)这样的认识论(epistemology)内容。本书总共有 1 000 多个词条,很多词条可互为参照。词条有 4 种不同的长度:约 2 500 字(A 水平)、约 1 000 字(B 水平)、约 500 字(C 水平)和约 50 字(D 水平)。一般来讲,词条越长,其阐述的论题越重要。

本《百科全书》的写作角度也是全方位的,许多专题都收录了来自不同视角的词条,既介绍使用得多的方法,也介绍使用得不多的方法。因为词条涉及的方法使用得多与不多并非是识别方法好坏或重要与否的标准。所有词条都由与之相关的主流学者撰写。

本《百科全书》的读者对象也是全方位的,除了社会科学的研究者和社会科学的统计学家之外,也应包括人类学、传播学、经济学、教育学、地理学、政治科学、心理学、公共卫生、公共政策、社会学、城市规划等各相关学科在读的本科生和研究生。不同层次与专业的研究者、实际工作者和学生都可从中各取所需,有所获益。

近年来,社会科学研究方法无论是定量方法还是定性方法,都取得了长足的进步,从分析模型的角度来看,定量分析在以下几个方向有了突破性进展:

1.缺失值处理。多重插值法(multiple implation)或者其他基于模型的方法的发展不仅大大提高了研究者的数据处理能力,而且改变了问卷设计的方式。

2.非线性关系模型日趋成熟。MARS 模型的出现使理论思考和数据分析成为一个互动的循环过程,定量分析不只是理论和假设的简单证伪过程,还是理论思维的一个重要组成部分。

3.测量技术突破了传统的四层次模式。潜类分析、广义线性模型的出现,大大提高了定量社会科学处理定类、定序数据和建模的能力。

4.测量模型有了很大进展。这方面的进展主要表现在两个方面:量表成分的分解;测量模型与因果模型或其他分析模型的结合,把测量误差明确引入分析中,充分评估它们对分析结果的影响。

5.潜变量技术得到广泛应用。“潜变量”使我们可以通过“显变量”来测量无法直接观察的理论概念,架起了连接理论和实践的桥梁。

6.多层次分析技术突飞猛进。这使我们可以同时考察多个层次上的问题,使宏观和微

观的联结获得了建模技术上的支持,宏观层次数据分析的结论有可能推论到微观层次。

7.社会网络模型应运而生。它将“关系数据”和“属性数据”做了区分,把分析重点从个体/群体等社会单元转移到这些社会单元之间的关系中。

8.系统动力学引进社会研究方法。事件史模型、事件数模型、历时分析、Cox 回归、时间序列分析、同期群分析、状态空间模型等的出现,使我们对社会过程在时间和空间上的属性的研究能力大大提高。

9.预测模型广泛应用。借助计算机技术的支持,神经网络、基因算法、人工智能、模式识别等数据挖掘技术有了长足发展,出现了许多拟合经验数据的预测模型。

10.社会应用研究中计算机模拟技术得到了广泛应用。计算机模拟已成为可供许多应用研究选择的替代方案。两种类型的模拟技术——基于计算机网络的模拟和基于概念模型系统的模拟都有很大发展,突破了历史、空间和伦理上的许多限制,使很多原来不可能的研究成为可能。

与此同时,定性分析也有了长足的进步,主要表现在以下几个方面:

1.研究素材趋于多样化。会话、交谈、电视、广播、文档、日记、叙事、自传等社会过程中自然产生的素材,甚至理论本身(理论的形式化)都开始进入定性分析的视野。所有这些资料,不仅可以以文本的格式存储,还可以以新型的多媒体介质,如图像、声音和视频格式存储,并作为原始的分析素材。

2.分析方法更加多样化。除了传统的源自语言学的方法之外,社会科学研究者也创造出各具特色的定性分析方法,把定性分析向更加系统、更加精确、更加严格、更加形式化的方向推进。

3.认识论基础更加多元化。实证主义也开始逐渐为定性分析所接纳,说明和阐释的关系由对抗逐渐演变为相互融合。

4.研究过程更加客观规范。定性分析开始引入定量分析中的“信度”“效度”“代表性”等概念,再加上传统的定性分析标准,通过编码和对比使得定性分析的过程更加规范,解释的方式更加客观,研究的结论更加可信。

5.研究过程更有效率。这主要应归功于计算机辅助定性数据分析(CAQDA)软件的出现和应用。据不完全统计,目前用于定性分析的软件已经有二十多种,这些定性分析的辅助系统,不仅使研究者从处理大量文字材料的繁复劳动中解脱出来,而且能够让研究者共享他们各自分析的细节,从而改变了定性分析的流程和研究团队之间的合作方式。

6.定性分析和定量分析的结合更加紧密。定量分析和定性分析方法相互融合有三个方面的取向:①将定性数据尽可能量化;②方法论多元化,即通过核心概念的测量模型,把定性分析和定量分析结合在一起;③在定量分析领域中更加注意“模型设定”问题,尝试使定性分析重返定量分析。

以上围绕社会科学研究方法所有的发展,都在这本《百科全书》中得到了充分的反映。同时,定量、定性两种方法的发展也催生了一种全新的方法论思想——总体研究设计(total reasearch design)思想,即在研究设计或评估时必须考虑与研究有关的所有方面和每一个环节,不要只是注意其中的某些方面和某些环节。对于研究方法的选择必须建立在对研究专题的全面和深入的了解之上,选择最合适的方法,无论是定量、定性抑或二者的结合,设计、测量和分析三者都融为一体。这本《百科全书》也是这种新型设计思想的体现。

重庆大学出版社最初找到我，提出让我翻译本书的想法时，我诚惶诚恐，深知自己的知识积累和外语水平都十分有限，难以担此重任。雷少波先生和诸位译者，特别是同事赵锋给了我很大的鼓励，使我回想起1980年代初，自己刚刚进入中国社会科学院社会学研究所的时候，时任所长费孝通先生在组织社会学恢复之后编写第一本社会学概论时讲的话。他说："当年编写社会学概论的都是学界的泰斗巨匠，如潘光旦、吴文藻、孙本文等先生，自己学浅才薄是没有资格的；但是现在社会学刚刚恢复，教学科研都迫切需要一本社会学概论，山中无老虎，猴子称大王，责任所系，只能勉为其难了。"前辈的榜样，更主要是社会科学研究方法发展的需要，最终使我鼓起勇气，承接了此项任务。

这样一部鸿篇巨著的翻译工作，只靠我一人之力是无法完成的，还需要一批像我这样敢做"山中之猴"的同仁一起努力。本《百科全书》的译者队伍以中国社会科学院社会学研究所社会调查和方法研究室的研究人员为骨干，还有来自高校和北京社会科学院的中青年教师及研究人员共9人。沈崇麟、赵锋和高勇为主译和审校。此外，出版社还专门从高校和研究机构聘请了10位词条审定专家。赵锋先生为总联络员，负责全部翻译工作的组织协调，对本书的完成做出了很大的贡献。

作为一部学术性译作，本《百科全书》译文遵循的原则可概括为"信、达、白"三字。与文艺作品的翻译一样，译者首先必须要准确理解原文的主旨，即所谓"信"；在这个基础上再将原文用译文确切地表达出来，即所谓"达"；但学术著作的译文与文艺著作也有所不同，译文不苛求文雅，却要力求简明直白，即所谓"白"。"信、达、白"三字是全书翻译的指导方针。对译文的具体处理，则在"信、达、白"的前提下，视实际情况，在权衡比较后作出定夺。所谓定夺是词条译者根据"信、达、白"方针作出的一种选择，不是定论，更不是定于一尊！具体的做法如下：

1.在音译和意译的选择上以意译为先。首先采用已有的正确译名，在已有译名不正确或可能会产生歧义时才采用新的译名，或直接采用英文。像Logit比较流行的译法是"逻辑特"。我们认为这种译法不妥，容易望文生义产生误解，以为它与逻辑有关。实际上这个词英文含义是很清楚的，Logit的中文意思就是取它的对数，故而这个词比较好的译法是"对数比"，同样，Probit应该是"概率比"，"Logistic"应该是"对数概率比"而不是"逻辑斯蒂"。所以本《百科全书》弃用了这种流行的译法，经再三斟酌，我们认为在这里还是直接采用英文原文，而不做画蛇添足的翻译，这样不仅不会引起不必要的误解和争论，而且更有利于学术交流。又如，Panel Data Analysis一词，流行的译法是"面板数据分析"，我们认为这显然是一种错误的译法，因为这里的Panel并非面板，而是指一群固定的调查对象，所以本书采用了"定群数据分析"这种译法，以纠正流行译法的错误。

2.在处理一词多名问题时，则在从众的前提下，由译者根据自己的考虑作出选择，如在Dependent Variable的众多译名"应变量""因变量""响应变量""依变量"中，选择了"因变量"这一译名。这个选择固然考虑到了从众，但更主要的是考虑到"因"这个字在中文中兼有"应""依"这两个含义，故而能够更确切地反映英文所包含的多重意义。但译者的选择并不意味着其他译法的错误，更不能将这种译法定为一尊。所以本《百科全书》并未将词条的其他各种正确译名作强行的统一，而是允许各种合理的译法存在，因为这样不仅能反映方法学的现状，也能引起读者的注意和思索，进而推动方法学的发展。

3.与此相关联，在遇到同一个词存在多种译法时，我们尽可能地采用与同义或近义词译

名不同的译名,如 approach 一词的译文是“路数”而不是用常用的“方法”。这样不仅可以区别 method 一词的常用译法,也能体现 approach 一词的本意。measure(ment)一词同时存在很多译法,如测量、测度、量度和度量等。我们则以译者所取译名为准,根据不同的语境采用了“测量”或“量度”。必须指出,这只是译者的选择,全无定名于一尊的想法。因而本《百科全书》在词条名以外的译文中,允许其他译名的存在,不作强行的统一,译文中采用何种译法,由译者视上下文内容、自己的习惯选择确定。

4.译文的外国人名则参考国内重要的工具书,采用“名从主人,约定从俗”的原则。外国地名、学术团体名、研究机构名由译者参照相关工具译出。人名、地名等凡专门工具中查不到的,则按通用译法译出。书名、期刊名、论文名等,有出版物的参考已出版物的译名译出;没有出版物的,则由译者自行定夺。

5.译文的其他细节处理:

(1)行文过程中以黑头括号的形式标注的英文意在提示读者,为进一步了解此处的内容可以查证相关的词条。比如,标注【Independent Variate】表示读者可以查阅词条“自变量”;标注【Field】表示读者可以查阅以“Field”为核心词汇的“田野研究”“田野笔记”等词条。

(2)为了方便读者根据目录进行检索,本书每一卷都生成了两个目录:第一个目录为英文原版的本卷目录,按照英文词条的首字母顺序编排;第二个目录为全书的总目录,该总目录根据词条的汉语拼音排序。

(3)本《百科全书》涉及众多的分支学科,同一概念各个学科的表达方式不尽相同,翻译过程中,遵从不同学科、作者、译者的使用习惯,只在必要的地方加上英文标注。

(4)文中重要的人名已生成汉英对照表放于书后供读者检索。

六年耕耘终于盼来了收获。本《百科全书》出版发行之时正是改革开放 40 年之际。回望自己的学术之路,恰与改革开放相伴。从某种意义上讲,社会科学研究方法在中国能有这样的发展,又何尝不是得益于改革开放的新政!当然只有改革开放的政策,没有一批为之努力奋斗,不断进取的践行者,也不可能有社会科学研究方法今天的发展。重庆大学出版社无愧是这批践行者中的佼佼者。他们以独到的学术眼光,投入了大量的人力、物力,组织出版“万卷方法”。自 2004 年至今,已引进相关专著一百余种,在我国社会科学界引起了很大的反响。而《百科全书》的出版,标志着“万卷方法”更上一层楼,达到了一个更高的水平。在这里我要向重庆大学出版社及负责和参与本书出版的团队表达诚挚的谢意。

本书翻译、编辑和出版的难度大大超出想象,虽然我们大家已经竭尽全力,但由于学识、能力所限,译文中一定还存在不少错误或偏颇,还望诸位读者不吝赐教,批评指正。

愿社会科学研究方法乘改革开放之风,借《百科全书》出版之力,在中国有更大的发展!

沈崇麟

# 前　言

《社会科学研究方法百科全书》是首次面世的有关社会科学研究方法方面的百科全书。它的这一特性至少部分解释了为什么我们要做这个项目。我们也相信诸如这样一本百科全书确实是必需的。什么是百科全书呢？在古希腊，这一词表示“全方位的教育”。这一参考集合即以社会科学研究者的方式给读者提供了一种全方位的教育。为什么这是必需的呢？因为每一个研究者，无论是专家，还是一般研究者，都会因为教学或研究的需要，而去查询某些方法学方面的知识点。而究竟应该到哪里去查询并非总是那么显而易见的。有时我们虽然知道应该到什么地方去寻找，但却不清楚应该找什么。本《百科全书》则几乎将所有的社会科学研究方法专题的权威论述集于一体。

一个抽样调查的研究者也许为了分析追踪的固定样本数据而想学习纵向分析方法；一个探讨住户劳动力分工问题的学生也许想学习如何使用时间日志；一个回归分析的专家可能专心致志于多维度量，以能更好地聚合形成某些量度；一个解构论者可能想要探讨一下福柯的话语分析；一个实验者可能希望对非实验者通常遇到的共线性问题有进一步的了解；一个女性主义学者可能力图对研究中电影的使用问题有更多了解；一位政治科学家可能一直都需要更新自己的实验设计；一位人类学家可能提出一个有关民族志的现实主义的问题；一位心理学家也许试图了解一个特定的关联量度；一个哲学家可能想阅读更多的社会科学定律；一位社会学家可能希望回顾一下不同的评估研究的路数。即使在某一个领域的那些专家，也可从本书中得到很多的知识。例如，在交互效应考察中，本《百科全书》有许多来自不同视角的词条，介绍使用得不多和使用得很多的方法，这些词条都是由知名的主流学者撰写的。

除了实践的研究者和社会科学的应用统计学家之外，本《百科全书》适用于更为广泛的读者。他们是一些什么样的读者呢？首先是世界各国大学的社会科学学科的学生，包括本科生和研究生。所有的主要分支——人类学、传播学、经济学、教育学、地理学、政治科学、心理学、公共卫生、公共政策、社会学、城市规划，都需要让本学科的学生阅读方法方面的文章和专著，使他们的知识更为全面。一个普通的本科生必须能读懂一张列联表，解释相关系数或对一个调查作出评价。本《百科全书》有些词条就与此有关，而其他一些诸如这样的词条，也完全可为这样的学生群体所接受。这些种类的词条既不需要专门的数学或统计学知识就能理解，也不需要学生做过实际的研究，便可抓住它们的基本要领。

当然，这些领域的研究生则必须超越简单的阅读理解，必须掌握这些方法的实际应用。这些词条中有许多是阐述“如何使用”那些方法的，如某些有关回归分析，或访谈技术的词条就是这样的。而本《百科全书》的有些词条则是比较高深的，这固然是对普通研究生的一个挑战，但他们也可从中受益。

除研究者和学生之外，本《百科全书》也会对受过高等教育，因工作性质（如杂志的编辑

或城市管理者),或个人喜好(如听懂媒体上的社会科学报告)而希望对社会科学的方法学有更多了解的读者有所裨益。许多词条都是以日常的英语撰写的,不要求读者具备专门的数学或统计学知识,因此完全可为非专业读者所接受。不仅如此,许多比较困难的词条都是介绍性的,因此是有可读性的。在说了前面那些话之后,我们要强调,本《百科全书》的词条没有一条"降低了强度"。相反,它们都是公认的专家的精心之作。这不是一本"菜谱"或高深莫测之作。在写作的风格方面,我们对作者的唯一要求是词条的长度,我们敦促作者在撰写词条时尽可能做到简明扼要。

我们给作者提供的词条撰写说明中包含了 4 种不同的长度:约 2 500 字(A 水平)、约 1 000 字(B 水平)、约 500 字(C 水平)和约 50 字(D 水平)。词条的长度最终取决于它们的重要性、内容涉及的广度,有时还要考虑论题的复杂程度。一般来讲,词条越长,其阐述的论题越重要。A 水平的词条如方差分析、贝叶斯推论、个案研究和回归。B 水平的词条也同样重要,但不需要更多的空间就能把概念解释清楚。B 水平的论题如历史方法、统计学、设定错误和混合设计。C 水平的词条通常用来处理相当专门的方法或论题,例如,曲线性、叙事访谈、囚徒困境和记录-核实研究等。A,B 和 C 水平的词条都以参考书目结尾,以引导读者进行进一步的阅读。另一方面,D 水平的词条都比较简短,主要是定义上的。

本《百科全书》总共有 1 000 多个词条,涵盖定量和定性方法及二者的结合。在本项目实施之初,我们打算提出一张穷尽所有可能方法的词汇的清单。我们查询了写作资料,也咨询了各界人士——国际学术界的杰出学者、方法学期刊的编委、本单位和其他单位的同事,排在最后但同样重要的是我们的研究生。随着这一进程的展开,有些词条被剔除了,有些词条作了编辑整理,有些词条则被加了进来。最终收入本《百科全书》的词条都是那些我们认为人们在阅读图书、杂志或报刊上的社会科学的研究成果时会遇到的词汇。秉承第一本由狄德罗和达朗贝尔于 1751 年编撰的百科全书的精神,我们的确试图"审视所有一切(Il faut tout examiner)",因为这与我们的事业有涉。

在词条清单基本确定之后,我们马上开始联系可能的撰稿人。我们为我们从社会科学界获得的广泛合作所感动。前辈学者和青年才俊都给了我们极大支持,并迅速给了我们回应。我们认为,这是一种工作很忙的研究者对这本《百科全书》的价值的一种量度。

除了撰稿人之外,本《百科全书》的编委、我们所在单位的同事和学生,还有其他很多人,都为《百科全书》的面世做了大量工作。SAGE 出版社在伦敦的高级编辑 Chris Rojek 提出了最初的构想;SAGE 出版社加利福尼亚的前高级编辑 C. Deborah Laughton 从一开始便是本《百科全书》构想的专注而不懈的支持者;Rolf Janke,SAGE 出版社的副总裁是来自出版社本部的主要支持者,他组织资源的能力和热情是本项目成功的重要保证。SAGE 出版社的技术编辑 Vincent Burns 则将网页、数据库和大量手写的数据收集到一起。此外,我们也不应当忘记制作编辑 Melanie Birdsall,是她耐心地带领我们审校了校样。最后,我们必须向我们的发展编辑 Eileen Haddox 致以衷心的感谢,是她细致入微的工作使我们的项目得以顺利进行。

米歇尔·刘易斯-伯克
艾伦·布里曼
廖福挺

(沈崇麟译校)

# 阅读导引

社会科学方法学可以说是在19世纪末进入我们视野的，那时埃米尔·涂尔干撰写了《社会学方法的规则》(*The Rules of Sociological Method*)一书，这是系统理解和分析社会现象的奠基之作。从那时起，它便一直在成长发展，在20世纪的最后几十年更是呈现出一种指数式发展之势。这种发展涉及方法学的专门知识和技能的广度及深度两个方面。应用于各种社会科学学科的方法，其涵盖范围及某些分析路数和技术的复杂与高深程度，都是在数十年之前想象不到的。例如，焦点小组和浓描实际上已经变成了社会科学研究者(不论其学科取向是什么)的通用语言。再举另外一个例子，处理缺失数据和无回答的方法取得了长足的进步，以致诸如这样的问题的专业工作坊一个接着一个，成为抽样调查数据分析者的最爱。

很自然，在21世纪之初，我们有必要全面盘点一下社会科学方法学的知识，但试图使我们的盘点包含社会科学中曾经使用过的每一种方法，是不可能和不切实际的。这本《百科全书》的主要目的是给我们的读者(他们可能是大专院校的本科生、研究生、学术研究者或实际应用者)遨游浩瀚的研究方法之海的一个引导。本《百科全书》对各种方法的目的、原理、发展和实际应用都作了简明扼要的介绍，总共有1 000多个词条，其中许多都对涉及的论题做了详尽的介绍，并包含了最新的发展，对初学者或有经验的研究者可能都有很大的用处。

为了达到这一目的，我们给读者提供了两种类型的词条：一些词条只含有一条定义，长度只有一段或两段。这些词条给了读者一个方法学词汇的简要解释。为了忠实于《百科全书》的行文形式，许多其他的词条都作为论题或论文来处理，长短不一，经常都附有实例，有时还有图表，对涉及的方法的性质、历史、应用和使用的含义进行讨论。这些词条的大多数也给出了建议的阅读书目和参考材料，以满足有进一步提高愿望的读者的需要。这些都是《百科全书》的精义所在，对那些想要对奇妙的研究方法世界进行不懈探索的读者而言都是很宝贵的。

需要撰写如此多样的专业文章，仅仅靠我们三位是无法完成的，幸运的是我们得到了编委和作者的大力支持，他们过人的才智，对一气呵成的、明了而又透彻的方法介绍作出了很大的贡献。有时，作者的行文风格如同新奇的小说，使得阐述的论题不仅能给读者以愉悦，而且发人深省。例如，达摩达尔·古雅拉提教授撰写的计量经济学的词条的文字给人以惊喜。他不只是仅仅介绍计量经济学家使用的方法和模型的类型，而不去论及其他的内容。古雅拉提带领我们，由平凡到不平凡地作了一次计量经济学的旅行。他用三个阐明计量经济学涉及的广泛范围的问题作为自己论述的开头，在这里他使用了简单而平常的路数，用引文释例完成了看似不能完成的、对计量经济学家的用武之地加以定义这一任务。然后，他两

次带领我们作研究(从经济理论到数据,从模型到分析),第一次是原理,第二次是实例。诸如这样的过程也是我们中的许多人平常反复讲述的过程,但我们几乎想不到在撰写《百科全书》的词条时用到它。古雅拉提用在经济学研究中求索的普通过程,给读者留下了不平常的、意义深远的影响,令读者去深思,而不只是记住计量经济学的方法和模型,以及基本目的。

词条的多样性是本《百科全书》的又一特点。以许多可能的对比中的一个为例,我们的词条中有些论及了哲学问题,例如,后结构主义,这样的内容似乎超出了本《百科全书》的范围,而还有一些词条则讨论了高级的统计技术,这似乎同样也不能被视为社会科学方法学的一部分。然而我们却认为二者都是必要的。一方面,所有的研究者都应该意识到,认识论问题影响着研究问题的性质并支撑研究过程方面的假定;另一方面,我们都需要了解分析定量数据的全套武器,这样才能在不同的场合使用最为恰当和稳健的技术。只有在我们对可供利用的各种选择——无论是认识论的、统计学的或其他什么知识都了如指掌时,作为社会研究者,我们才能完整地开发出精巧的社会科学研究的方法。

本《百科全书》的许多词条都有精当的例子。将各种来源适用于不同研究目的的词条汇编成一册百科全书,我们,本《百科全书》的编辑者,固然希望读者诸公将会钟情于我们社会科学方法学的博大精深,但我们更希望读者诸公能从这些方法学的专业知识中有所斩获,推动自己的研究更上层楼。

米歇尔·刘易斯-伯克<br>艾伦·布里曼<br>廖福挺

(沈崇麟译校)

# 读者指南

## 变量分析

协方差分析
方差分析
主效应
方差分析模型 Ⅰ
方差分析模型 Ⅱ
方差分析模型 Ⅲ
单因素方差分析
双因素方差分析

## 关联与相关

关联
关联模型
不对称量度
双列相关
典型相关分析
相关
对应分析
组内相关
多重相关
部分相关
偏相关
皮尔逊相关系数
系数
半偏相关
简单相关（回归）
斯皮尔曼相关系数
关联强度
对称量度

## 基本定性研究

自传
生活史方法
生活故事访谈
定性内容分析
定性资料管理
定性研究
定量和定性研究论争
定性资料的二次分析

## 基础统计

备择假设
平均数
条形图
钟形曲线
双峰
个案
因果建模
格
协方差
累计频次图
数据
因变量
离中趋势
探索性数据分析
*F* 比率
频数分布
直方图
假设
自变量

## 因果建模



## 话语/谈话分析



## 计量经济学



## 认识论



## 民族志

**评估**



**事件史分析**



**实验设计**



**因子分析**



**女权主义方法论**



**广义线性模型**



**历史/比较**



**定性研究中的访谈**

## 多层次分析

情境效应
相依观察值
固定效应模型
混合效应模型
多层次分析
非参数随机效应模型
随机系数模型
随机效应模型

## 多元回归

调整的 $R$ 方
最佳线性无偏估计量
Beta($\beta$)
广义最小二乘
异方差性
交互效应
误设
多重共线性
多元回归分析
非加性
$R$ 方
回归
回归诊断
设定
估计标准误

## 定性资料分析

分析式归纳
CAQDAS
持续比较法
扎根理论
原生编码
备忘录,撰写备忘录
不符案例
定性内容分析

## 定性研究的抽样

目的抽样
定性研究的抽样
滚雪球抽样
理论抽样

## 调查抽样

多阶抽样
配额样本
随机抽样
代表性样本
抽样
抽样误差
分层样本
系统抽样

## 度量

态度测量
双极量表
维度
双重度量
哥特曼度量
指数
李克特量表
多维度量(MDS)
最优度量
量表
度量
语义差异量表
瑟斯通度量

## 显著性检验

Alpha($\alpha$),检验的显著性水平
置信区间
显著性水平
单尾检验
检验的统计效力

## 简单回归



## 调查设计



## 时间序列

# 本卷目录

## A

## B

# C

# D

## E

# F

# 总目录

（按汉语拼音排序）

## A



## B

## C

## D

## E



## F

## G

## H



## J

**K**

## L



## M

## N



## O



## P

## Q



## R

## T

## V



## W



## X

## Y

# Z

# A

## 常情诠释(Abduction)

常情诠释是以社会行动者的日常行动、使用的语言及其意义为根据建立的描述和解释逻辑。常情诠释指的是一个过程,即从社会行动者对其生活方式的描述向对其社会生活进行技术性的社会科学式描述转变的过程。其中包含了两个阶段:(a)描述这些行动及其意义,以及(b)在对当前问题的理解或解释的基础上发展出类别和概念。常情诠释与诠释主义【Interpretivism】存在关联。

常情诠释的逻辑在很大程度上通过对人们不言而喻的共同知识以及作为其行动指南的象征意义、动机和规则的揭示,来发现人们为什么那样做。共同知识主要是难以言明(unarticulated)的背景知识,在社会行动者与他人的互动中不断被使用和调整,在社会行动者共同的生活过程中被生产和再生产出来。它是日常的信仰和实践——世俗平凡且理所当然,社会研究者必须把握和阐明它,以此来为人们的社会行动提供理解。

常情诠释的概念在哲学和社会科学中的应用很有限。哲学家皮尔士(Peirce,1931,1934)和社会学家大卫·维勒(Willer,1967)曾使用过这个概念,但其意义跟这里有所不同;我们这里所使用的概念来自布莱基(Blaikie,1993)。尽管舒茨(Schütz,1963)没有使用常情诠释的概念,但他构想出了这一逻辑,而后杰克·道格拉斯、约翰·雷克斯,尤其是吉登斯(Giddens,1976)都使用了这一逻辑(对这些作者的评述请参见Blaikie,1993,pp. 163-168,176-193)。

舒茨、道格拉斯、雷克斯和吉登斯的共同之处在于,他们都相信社会科学对社会世界的解释必须源自社会行动者对社会科学家感兴趣的方面所能给予的解释,至少最初必须如此。不过他们在如何处理这些解释这个问题上意见各不相同。有人认为报道日常的解释就是理解社会生活的可能或必要的全部;而另一些人却准备把这些解释转化成社会科学对某个特殊社会群体(社区或社会)的生活方式的描述,但他们坚持这些描述应与社会行动者的语言紧密相关。这些描述为他们带来了另外两种可能:第一种是拿某种既有的理论或观点套在上面,从而提供了一种社会科学的解释,或对那种生活方式的批评。第二种是以日常解释中发展出来的一些理想类型为材料,来形成某种解释。后面这两种可能都存在一些争议——就第一种而言,争议主要围绕在对别的生活方式的批评是否正当这一问题上,而对第二种,争议则主要围绕在为了形成一种理论,究竟还有什么其他东西能够或者应该加到日常的解释之上这个问题上。

诠释主义的社会科学家使用的多种研究方法都包含着某种版本的常情诠释。尽

管如此，但他们却很少明确阐述其逻辑。运用常情诠释的适当研究方法还在不断开发之中。

——Norman Blaikie

（杨可译　沈崇麟校）

## 参考文献


Blaikie, N. (1993). *Approaches to social enquiry.* Cambridge, UK: Polity.

Giddens, A. (1976). *New rules of sociological method.* London: Hutchinson.

Peirce, C. S. (1931). *Collected papers* (Vols. 1-2; C. Hartshorne & P. Weiss, Eds.). Cambridge, MA: Harvard University Press.

Peirce, C. S. (1934). *Collected papers* (Vols. 5-6, C. Hartshorne & P. Weiss, Eds.). Cambridge, MA: Harvard University Press.

Schütz, A. (1963). Concept and theory formation in the social sciences. In M. A. Natanson (Ed.), *Philosophy of the social sciences* (pp. 231-249). New York: Random House.

Willer, D. (1967). *Scientific sociology: Theory and method.* Englewood Cliffs, NJ: Prentice Hall.


# 调查通道(Access)

所谓调查通道是指研究者正在寻找的一种入口，通过这个入口，研究者可以进入一个可以发掘和产生基本研究数据的地方。就研究者而言，其研究目的都在于分析那些定性或定量的基本数据，因而从这个意义上讲，寻找进入通道是研究过程中不可或缺的步骤。正因为这样，保证调查通道的畅通便变成了研究者所应掌握的一个必要研究技巧。但是在实际寻找调查通道的过程中，我们很少能采用系统的首尾一贯的考察方法，因为此间研究者常常需要采用一种随机应变的、因地因时制宜的方法。在考察那些已为研究者使用过的一系列用来确保调查通道完全畅通的策略时，首先考虑一下研究者寻找调查通道来做什么这一问题，不失为一种明智之举。

## 通过通道去找什么和通过什么人来找

调查通道常常被看成是一种叫作封闭的和开放的/公共的环境(closed and open/public settings)(Bell, 1969)。所谓封闭的环境，如商号、医院、教堂和它们的档案材料等。而开放/公共的环境的性质就不像封闭的那么正式。深入街角或某地音乐会现场开展的研究，便是开放/公共环境研究的一个例子。找到进入封闭或开放环境的调查通道的关键在于与看门人(gatekeepers)进行谈判。所谓看门人，不言而喻，是指那些有能力打开或关闭那扇研究者正在寻找的通往他们想去的那些环境的通道大门的人。在寻找进入通道时，最为主要的一步是确定看门人。在正式场合，封闭环境看门人通常都是那些身居高位的人，例如高级主管。而在不那么正式的开放环境中，看门人则不那么容易确定，不过他们通常是在那些环境中有地位和受人尊敬的人。有些研究者在试图确保进入通道畅通无阻时，自己直接去见看门人；而有些研究者，则先通过第三者，如该环境中研究的“拥护者”或“赞助者”，借助这些人的帮助，得到看门人的应允，打开进入通道的大门。

## 获取调查通道的过程，就是信任形成和谈判协商的过程

研究者争取得到那些赞助者、主要的看门人或分散的个人信任的能力，是确保调查通道畅通无阻的关键。获取调查通道可看

作一个与相当广泛的社会问题有关的信任建立问题的一个特例。因而我们可参阅那些有关信任建立问题的文献，这将有助于我们更透彻和更为系统地理解确保调查通道畅通的过程。科尔钦斯基（Korczynski，2000）认为，信任的四个主要方面有可能构成一种人际交流的基础。在协商谈判的过程中研究者使用最为普遍的策略大致有以下几类：

- *基于人际关系的信任*。研究者常常将自己认识的朋友和熟人作为可以信赖的获取调查通道的途径。这些朋友和熟人或可作为他们与赞助者取得联系的通道，或有可能直接作为调查赞助者，或研究环境中的实际的看门人。
- *基于对他方内在规范的了解的信任*。研究者有可能对某个"合适的"看门人或个人给出一些表示来取得他们的信任。例如，通过在持有某一特定的道德、社会，或政治观点的报纸或杂志上发公告的方法来寻找通向回答人的调查通道。在公告中比较含蓄地邀请那些持有与报纸杂志类似的道德规范和价值观的读者与研究者联系。
- *基于以交流为中心的某种激励或治理结构的信任*。研究者若能强调自己绝不会辜负当事人的信任，这一点将有助于信任的产生。例如，研究者可以给看门人否决研究发现发表的权力（尽管这样一种做法可能会将研究者置于一种潜在的危险之中）。
- *基于某些系统或机构的信任*。研究者可以自报家门，设法使看门人或其他个人对自己产生信任。例如，研究者可以先宣称他工作的大学声名如何卓著，进而介绍自己的大学近来所做的有关研究。

获取调查通道的过程应该被看作一个协商谈判的过程（Horn，1996）。研究者一般都希望尽可能拓宽和加深进入某一环境的调查通道，而看门人和其他有关人等则担心建立诸如这样的通道可能花费的代价。这样一些潜在的利益冲突，在面对面的会谈中尽管没有明言，但彼此都是心知肚明的。在这样一些心照不宣的协商谈判中，研究者一般都会给看门人或有关人等一些承诺或好处，保证将要给予的调查通道的安全。而在比较正式的场合，就需要将作出的决定写成书面报告的形式。

信任的产生和协商谈判似乎是一个高度随情境而定的过程。而这恰恰在一定程度上说明，在讨论研究的调查通道问题时，为什么研究者总是采用那些强调视情境而定的策略。迄今为止，几乎没有人尝试更为抽象地理解确保研究通道畅通的过程。

## 调查通道的获取是一个不间断的过程

在很多定性研究的研究者看来，协商谈判调查通道贯穿整个研究过程。例如，一个想要对某一正式团体的成员进行个别访谈的研究者，必须依次与每一个访谈对象进行协商谈判，一一获取每一个访谈对象的调查通道，而不可能只是依靠从一个非正式的看门人那里轻而易举地得到的予以合作的含蓄的应允。不仅如此，在正式场合，尽管回答人觉得自己有责任接受研究者的访谈，但在实际回答问题的过程中，仍然会有所保留。研究者需要在不同情况下，在持续不断的协商谈判过程中产生信任。

## 调查通道和反应性

所谓反应性是指研究者影响研究环境的能力。获取调查通道采取的协商谈判的方式会对这种反应性有显著的影响。例如，在一个组织研究中，如果只是与高层的看门人，而没有与那些包括在该研究中的底层员工形成信任的话，那么这种确保调查通道畅通的模式就可能对研究产生严重偏倚。诚如伯吉斯（Burgess，1984）所言："调查通道会对研究者实际得到的数据的信度和效度

产生影响”(p. 45)。

——Marek Korczynski

(沈崇麟译校)

## 参考文献

Bell, C. (1969). A note on participant observation. *Sociology*, *3*, 417-418.

Burgess, R. (1984). *In the field*. London: Allen & Unwin.

Horn, R. (1996). Negotiating research access to organizations. *Psychologist*, *9*(12), 551-554.

Korczynski, M. (2000). The political economy of trust. *Journal of Management Studies*, *37*(1), 1-22.

# 陈情(Accounts)

陈情是一段谈话,用以说明自己的来意,或拉近某些人与另一个人的关系。在一篇颇有影响的论文中,社会学家斯科特和莱曼(Scott & Lyman,1968)强调,如何谈话是我们借以处理与他人关系的基本手段。陈情谈话的目的是设法弥补我们与他人之间存在的问题,而更为广义地说,陈情谈话的目的在于使行动合法化。陈情有两种重要的方式:(a)为行动的陈情(accounts *for* action),(b)对行动的陈情(accounts *of* action)。

为行动的陈情是一种对不恰当的行为作解释的陈述。所谓的不恰当行为也许是一种微不足道的行动,如忘记事先用电话预约、迟到,或开了一个不合时宜的玩笑。不恰当行为也可能是一些重大过失,如未经许可拿人家的钱、婚外情,或在一些重大的事情上撒谎。陈情也可能是一个人为另一个人作的解释性陈述,例如员工在为自己的老板的粗鲁行为道歉时,常常会说老板的这一行为有点反常。为不当的行为陈情经常采用道歉或辩护的方式。在道歉时,当事人先承认自己的确错了,但随后则解释道,错的责任并不完全在自己。如学生对他的老师说“对不起,作业交迟了,因为昨天我的硬盘坏了”,便是这样一种方式的道歉。辩护性陈情则不同,当事人承认对自己的行为应负全部责任,但却不认为行为有什么不对。这方面的例子譬如,一个十几岁的孩子在听到他的父母说他上学要迟到了的申斥时,说道:“我已经是大人了,可以自己作决定了,你们没有权利告诉我要做什么。”道歉和辩护是两种不同的陈情方式,但有时它们却可以配对使用。譬如一个人可能对他的合伙人说:“对不起,我刚才对你嚷嚷了。我最近一直处在高度的压力下,况且当时一个电话会议正在进行中,你想我又能怎么样呢?”这个人在道歉和辩护的同时也赔了礼(承认自己行为不当,并表示为此感到懊悔)。

第二种使用更为广泛的陈情方式是对行为的陈情。对行为的陈情(给出行为的理由)是日常会话具有的一个普遍的性质。每当人们在为自己的行为作解释时,他们便是在进行陈情。一个妇女在告诉她的朋友,她为什么嫁给了蒂姆的时候,她就是在陈情。某人的同事告诉人们他为什么不再吃红肉,或一个丈夫在向妻子解释,他们为什么应该去葡萄牙而不是希腊,也都是在陈情。行为陈情常常以一种细说的方式出现。无论为行为还是对行为的陈情都有一种不太明显的性质,那就是它们都在告诉我们一个群体或一种文化的一些情况。在人们对他们的

行为给出解释(或想法)时,他们便是在将自己的行为作为一种选择。注意,人们并不是在对那些他们认为是自然而然,或必然的事情作出解释。因而,陈情便成了一种进入某一种文化的窗口。通过对陈情者、陈情内容和陈情环境的审视,我们得以了解合理行为背后的文化信仰。

——Karen Tracy

(沈崇麟译校)

## 参考文献


Buttny, R., & Morris, G. H. (2001). Accounting. In W. P. Robinson & H. Giles (Eds.), *The new handbook of language and social psychology* (pp. 285-301). Chichester, England: Wiley.

Scott, M. B., & Lyman, S. (1968). Accounts. *American Sociological Review*, *33*, 46-62.

Tracy, K. (2002). *Everyday talk: Building and reflecting identities*. New York: Guilford.


# 行动研究(Action Research)

行动研究是一种将当地人作为研究者合作伙伴的研究策略。行动研究的特点是合作、互教和为改变而行动。这种研究路数通过对研究环境中语境因素(contextual factors)的识别来提高研究的效度,而这一点常为那些更具结构性的研究路数所忽略。行动研究对性别、经济地位、能力,以及那些可能对研究合作伙伴、结果和研究社群有影响的文化因素比较敏感。

道德规范是行动研究最主要的主题:研究者建立一种有某种意图的道德立场。这种立场对研究进程、哲学框架和评价方法是有意义的。为了认定那些将要引导研究进程的公共价值,通常伦理原则最初在所有的合作伙伴间是可以商讨的。这些原则可能包括将商议贯穿整个研究过程、评价各种知识、互相尊重,以及商谈研究的问题、方法、宣传和后续行动。

行动研究可能会涉及整个研究项目,如研究团队的确认和招聘、研究问题的定义、数据收集方法的选择、数据的分析和解释、研究结果的宣传、研究结果和进程的评估、后续项目的设计以及各种服务或政策的改进等。尽管合作的程度和社区卷入的程度因研究项目而异,但根据参与者的需要或愿望的不同,行动研究在许多方面仍然可以是“有所为”的。所有合作研究者,包括社区成员在研究过程中都起着积极的作用,研究过程本身就是一个培养所有研究合作伙伴能力的培训班,在研究团队内,所有成员分享当地的知识和研究技巧,而研究成果便是一个项目的建立,或政策的制订。理想化地说,行动研究对所有的参与者都是一种改变。

## 历史

行动研究发端于库尔特·卢因(Lewin, 1935,1959)的研究工作,继而在社会学、人类学、教育和组织理论的论述中得到了发展。它也发端于1970年代的发展理论,发展理论提倡所有研究合作伙伴的平等,以及知识、技巧和经验的互补。伴随着人权运动对传统科学研究方法的挑战,合作研究的新模式也应运而生。这些新模式有着各种名称:行动研究、参与研究(PAR)、合作研究及开放的、授权的或以社区为基础的研究。在这些名目繁多的研究之间有相当多的重叠。所有这些模式有一个共同之处,那就是它们都承诺研究的目的在于将传统的研究范式予以扩充,以使它能包括那些研究产生的最

大的影响，进而引起正面的社会变迁，这样，在研究过程中，“研究对象”就可以成为积极的参与者。其中大多数模式都包含有一些能力的培养或训练要素。这些要素有助于研究得到社区许可，提高宣传过程的适当性，并使行动研究成果得以持续。

## 实例：加拿大肺结核防治计划

加拿大肺结核计划是一项三年计划。该计划的目的在于在那些最容易传染肺结核的人群——国外出生和土著人口中，确认影响肺结核防治和治疗的社会文化因素。它是由阿尔伯达省(province of Alberta)的一个从事肺结核工作的护士发起的。这一研究需求得到了肺结核感染者群体领导者的响应。该研究一开始建立了一个社区咨询委员会(CAC)，并协商确定了一系列指导原则。这些进程有助于研究团队的各种成员——包括来自社区、政府和学术机构的研究合伙人——之间的信任建立和团队建设。然后CAC和行政工作人员一起，从四个土著和六个国外出生的社区中各招聘两个代表。这20位社区研究合作人接受了行动研究原则、访谈技巧和定性数据分析方法的培训。他们执行了他们所在社区的四个群体的访谈。访谈的对象包括肺结核患者、接受肺结核预防者、拒绝肺结核预防者，以及在原籍国或土著原住地有更久远的肺结核家庭史者。然后，社区研究合作者、CAC和学术机构的工作人员一起分析，用手工和电子技术同时分析访谈数据。这种多种方法和多个调查者的分析策略最大限度地提高了数据和研究发现的结果的效度。

行动研究过程对合作团队是很宝贵的这一点已经十分清楚，因为它使团队的成员能学会尊重其他每一成员了解的肺结核传染和治疗的方式方法。不仅如此，研究结果还进一步确认，现行医疗卫生系统固然对处在活动期的肺结核的治疗已经是非常有效的，但在肺结核高危社区的防治工作仍然还是非常薄弱的。此外，专业医务人员在筛查肺结核患者的时候经常发生遗漏，因为在西方国家的一般人口中，肺结核患者并不多见。研究结果要经过CAC和社区的合作研究者复查，以确认它们是否符合既定的道德原则。在此之后，研究团队将提出一份宣传计划和后续的行动计划，其内容包括视频宣讲材料、建立社区护士培训岗位以及各种语言文本的感染肺结核宣传单等。

## 行动研究的局限性

行动研究路数可能因为不够严格和有调查偏倚而受到批评。然而不只是行动研究，实际上所有的研究都可能遇到这样的问题。无论采用什么样的研究路数或方法，只有在我们对这些问题保持高度的警觉时，才能保证研究过程和结果的高质量。行动研究遇到的其他问题包括研究项目耗时较长、长期的承诺和资助、使多个和/或没有经验的研究人员协同一致、说服资助者支持这样的研究路数、持有很强理论色彩的研究框架、减少参与者的流失等。

## 行动研究评估

行动研究的复杂性决定了它需要一个精细的文档化和分析式的审查跟踪。让所有的利益涉及方参与设计和实施的全过程的评价方式，可以增加数据和评价过程本身的效度(Reason & Bradbury, 2001)。只有在将整个评价整合成整个研究计划的一个常规活动时，这些评价才会最有效。在对方法论和研究计划上的目标作量度时，若能将量度与研究背景相关联，那么研究的严格性同样也会因此而有所提高(Gibson, Gibson, & Macaulay, 2001)。在对涉及最初设定的目

标进行测量和对协商确定的道德规范的一致性进行评价的同时,包容行动研究的动态性是评价面临的一个重大挑战。

## 结果和行动

行动研究经历的过程,包括互相学习和通过培训培养能力等活动,是与具体的研究结果同样宝贵的。因此它的宣传策略应该包括研究中的各种创新机制,例如那些可以为社区成员和学者使用的海报,以及那些适宜于社区使用的媒体,如收音机、电视,时事通讯、镇民会议和学术文章(这些文章在发表之前,应经所有合作研究伙伴审阅,以免未经授权的内容发表,或在不经意间对社区的情形误传,或不客观的描述)。因为行动研究的教育目标、谈话或海报可以在多种场合以多种语言实现,所以它们本身既可以作为一种培训的手段,也可以作为一种宣传的策略。行动研究过程可以产生高效的联盟,建立起跨越存在于当地社区、学术机构、政府部门和各种组织之间的陈年鸿沟的桥梁。

——Nancy Gibson

(沈崇麟译校)

## 参考文献

Cooke, B., & Kothari, U. (Eds.). (2001). *Participation: The new tyranny?* London: Zed.

Gibson, N., Gibson, G., & Macaulay, A. C. (2001). Community-based research: Negotiating agendas and evaluating outcomes. In J. Morse, J. Swanson, & A. J. Kuzel (Eds.), *The nature of qualitative evidence* (pp. 160-182). Thousand Oaks, CA: Sage.

Lewin, K. (1959). *A dynamic theory of personality: Selected papers* (D. K. Adams & K. E. Zener, Trans.). New York: McGraw-Hill. (Original work published 1935.)

Reason, P., & Bradbury, H. (Eds.). (2001). *Action Research: Participative inquiry and practice.* London: Sage.

Stringer, E. T. (1996). *Action Research: A handbook for practitioners.* Thousand Oaks, CA: Sage.

# 积极访谈(Active Interview)

"积极访谈"这个术语并非指一种特定类型的访谈,而更多表明的是一种访谈过程中的具体方向。它表达了这样一种观点,即所有的访谈都不可避免地具有主动性、互动性和对话性特点。任何访谈中那些动态的、沟通性的意外都会实实在在地刺激被访者的观点。每个访谈都不仅是对信息的发现和传递,还是一个建构信息的机会(Holstein & Gubrium, 1995)。

访谈【Interviewing】总是对话式的,与高度结构化、标准化的问卷调查不同,访谈是更为自由和非正式的信息交流。访谈中出现的叙述可以简略得像由强迫选择测验组成的问卷结果,也可以细致得像口述生活史,但通过访谈参与者的谈话,这些访谈叙述本质上都是具有主动性和建构性的。

尽管大部分研究者都承认访谈的互动特征,但关于访谈技术的文献资料却倾向于关注如何使访谈受到控制。访谈的指导者,特别是致力于使调查标准化的研究者,首要的关注点是让有效的、可靠的信息流最大化,同时让对被访者所知信息的歪曲最小化。从这个角度看,标准化的访谈对话是传导信息的渠道,而那些非标准化的互动代表了混杂信息的潜在来源,而在调查中,混杂信息应当最大限度地减少。

另一方面，查尔斯·布里格斯（Briggs，1986）指出，访谈必然会从根本上形塑谈话的形式和内容。亚伦·奇库雷尔（Cicourel，1974）则进一步指出，访谈会将对事实的特定理解和传播方式强加于被访者的回答之中。问题在于，访谈者会深深地和不可避免地被牵连到那些表面上属于被访者的意义创造。因而，访谈中的双方都必然是积极主动的。意义并不仅仅通过具有倾向性的提问所引出，也不是简单地通过被访者的回答来传导，意义以交流的方式由访谈各方共同构建。被访者不仅是经验信息的存放库，而且也在与访谈者的合作过程中成了知识的建构者（Holstein & Gubrium，1995）。

对于“访谈具有主动性”这一特点的感知，会导致一个重要问题，那就是使用传统研究路径所预设的方式进行信息收集和分析有多大的可能性（Gubrium & Holstein，1997）。它也促使访谈者更关注在访谈过程中获取信息和理解的方式，而不再像传统路径那样只将访谈认定为研究案例。换言之，理解在访谈中意义创造的过程是如何展开的，与理解实实在在的问答内容具有同等重要的地位。如何访谈这个问题，当然不仅指访谈技巧，还指知识生产中的互动性、叙述性程序。访谈内容这个问题，则包含了访谈的主题指引、提问问题的内容，以及被访者回答的真实信息。对于意义建构中如何访谈和访谈内容这两方面的双重兴趣，始终与重视访谈过程中的建构主动性（constitutive activeness）齐头并进。

——James A. Holstein
Jaber F. Gubrium
（郑晓娟译　沈崇麟校）

### 参考文献

Briggs, C. (1986). *Learning how to ask.* Cambridge, UK: Cambridge University Press.

Cicourel, A. V. (1974). *Theory and method in a critique of Argentine fertility.* New York: John Wiley.

Gubrium, J. F., & Holstein, J. A. (1997). *The new language of qualitative method.* New York: Oxford University Press.

Holstein, J. A., & Gubrium, J. F. (1995). *The active interview.* Thousand Oaks, CA: Sage.

## 调整的 *R* 方（Adjusted *R*-Squared）

参见拟合优度量度【Goodness-of-Fit Measures】；多重相关【Multiple Correlation】；*R* 方【*R*-Squared】。

## 调整（Adjustment）

调整是一种（数据）处理分析方法。这种方法的目的在于将一个变量中那些并非我们主要感兴趣的预期效应量，从我们主要感兴趣的关系中去除。这一目的可借助于若干种已为人们普遍接受的程序来实现。

调整的需要发端于研究者观察到变量之间那些引起他们倾情关注的关系中，也含有其他一些他们并不直接感兴趣的变量的关系。种族特性与艾滋病（AIDS）关系也许

会使研究者对之产生很大的兴趣，但我们一定也知道，因这种疾病引起的死亡也会受到艾滋病人其他特征的影响（例如教育、文化适应性和卫生保健等）。研究者面对的基本问题是，种族特性与艾滋病引起的死亡的确有关系，还是根本就没有任何关系。研究者可能面对的第二种情况是，种族特性只不过是寄生其上那些其他真正决定着累计死亡率的变量的组合的代理载体而已。因此解释这些其他的解释变量（或协变量）和种族变量以及艾滋病死亡之间的关系这一问题就变得非常重要。用于调整一个变量的效应的三种分析方法是：(a)控制，(b)分层，(c)回归分析。在用控制方法时，一个其效应要进行调整的变量（调整器）是不允许以某种重要的方式变化的。对于二分变量而言，这一点是十分容易做到的。例如，我们认为刑期和犯人在判刑时的年龄的关系会受到他们的性别的影响，那么我们就可通过只研究男性犯人的刑期和年龄之间的关系来去除这一效应。借助将性别固定在一个，且只固定在一个值上这一方法便可将性别的效应抵消。

分层处理法增加了一个层。它使变量调整过程的实施程序变得更为复杂。特别需要注意的是，分层法使用了一种重复控制技术，分析的是在一种特定的背景下我们主要感兴趣的变量之间的关系。在这种背景下，调整器先被固定在某一个水平，然后再被固定在另一个水平。在调整器变量水平数比较少的时候，这种方法是十分灵便的。例如，在前面那个例子中，我们可以在(a)只在男性犯人中，或(b)只在女性犯人中来评估刑期和年龄之间的关系。这样的评估使我们得以考察性别在修正刑期和年龄之间的关系中的作用。统计方法使我们可以在每一个层中考察刑期和年龄之间的关系的同时，跨所有的层从整体上确认我们主要感兴趣的那两个变量之间的关系。如果调整变量并非二分的，而是连续的，那么我们可以以调整器的一系列值为依据，将它分成若干个不同的层。

虽然用回归分析来进行调整的方法多少会复杂一些，但是若使用得当，则能圆满地处理一些复杂的调整问题。在对那些调查者认为因某种关系而连接在一起的变量之间的关联强度进行考察时，回归分析本身就是一种重要的工具。不过，这样一种回归模型必须精心设计，务求精确。我们先从回归分析中一种简单易懂的回归模型——简单直线模型开始。假设我们收集到了 $n$ 对观察值($x_i$, $y_i$)，$i=1\sim n$。调查者的目的在于证明具有不同的 $x$ 值个体，会有不同的 $y$ 值，其模式可用某种直线关系作最好的预测。我们将 $y$ 对 $x$ 的回归写作下式：

$$E[y_i] = \beta_0 + \beta_1 x_i \qquad (1)$$

例如，$y$ 可能是犯人被判的刑期，而 $x$ 则可能是犯人在判决时的年龄。而调查者则认为，在决定刑期的长短时，犯人的社会经济地位（Socioeconomic Status, SES）也起着一定的作用。在这样的设定中，令 SES 保持不变可能并非一种可资利用的路数，因为这是一个变化范围很大的连续变量。相反，我们可以来考虑一下下面那样的模型，在该模型中，犯人的刑期同时回归于年龄和 SES 两个变量：

$$E[\text{刑期}] = \beta_0 + \beta_1(\text{年龄}) + \beta_2(\text{社会经济地位}) \qquad (2)$$

普遍使用的用于估计参数 $\beta_0$，$\beta_1$ 和 $\beta_2$ 的最小平方（亦称最小二乘）估计法可从许多统计软件包中找到。我们感兴趣的参数的估计值 $\beta_1$ 可以通过三个回归求得。首先做因变量刑期对调整器 SES 的回归。这一回归遗留的问题是刑期变量中 SES 无法解释的那一部分。这无法解释的部分便是刑期的残差。接下来，再做年龄对于调整器

SES 的回归,与此同时计算年龄的残差。最后,再做刑期残差对年龄残差的回归。这最后做的回归结果便会产生方程(2)中 $\beta_1$ 的估计值。

在大多数回归分析的软件中,这些计算都是自动进行的。这个回归模型得到了一个刑期和年龄之间关系的估计值,在这一估计值中,SES 的影响已经被"去除"。需要特别指出的是,这种回归程序先隔离并确定了刑期和调整器 SES 之间的关系,再将调整器的影响去除,然后将这一程序重复使用于年龄。最后,程序做了刑期的残差(即在去除了 SES 的影响之后剩下的刑期)对年龄的回归。因而在回归分析中,调整有着不同于对调整器进行控制或分层的特别含义。在回归分析中,调整意味着隔离、识别和除去调整变量的影响。在解释变量是离散的时候,方差分析【Analysis of Variance, ANOVA】便可得到这样的结果。在解释变量由离散和连续变量混合组成时,协方差分析【Analysis of Covariance, ANCOVA】则会产生相同的结果。

在回归模型中有一个以上的调整变量时,这一处理方法同样也可采用。这时,程序给我们提供的是在其他一组变量同时进行调整之后的我们所感兴趣的变量之间的关联强度的估计值。

——Lemuel A. Moyé

(沈崇麟译校)

## 参考文献

Deming, W. E. (1997). *Statistical adjustment of data*. New York: Dover.

Kleinbaum, D. G., Kupper, L. L., & Morgensten, H. (1982). *Epidemiologic research: Principles and quantitative methods*. New York: Van Nostrand Reinhold.

Neter, J., Wasserman, W., & Kutner, M. (1990). *Applied linear statistical models* (3rd ed.). Homewood, IL: Irwin.

# 聚集(Aggregation)

聚集是把不同于最终目标水平的较低水平的数据汇集起来的过程。数据之所以会聚集,是因为数据在群体水平更容易被收集到。例如,一般讲,犯罪统计数据更容易在市中心,而非该市的每一个市民处收集到。然后我们便可试图确定不同聚集之间的暴露-反应关系(exposure-response relationships)。所谓区位谬误【Ecological Fallacy】,是指这样一种谬误——将一个来自个体的群体或聚集的结论,应用于这些聚集内的个体。因为每个聚集中的个体是彼此不同的,所以有关某个聚集的结论是不可以用于它的每一个成员的个体的。

——Lemuel A. Moyé

(沈崇麟译校)

*也可参见区位谬误【Ecological Fallacy】。

# AIC

参见拟合优度量度【Goodness-of-Fit Measures】。

# 算法(Algorithm)

算法是计算机科学用于解决某一问题方法的术语，也是对计算机发出的解决这一问题的指令。虽然算法问题主要是计算机科学研究的对象,但对于社会科学中数据分析的实践也是非常重要的,因为它也常要在各种统计程序中使用。

算法可被看作任何一种解决任务的一步接一步的程序。设想有五张面朝下放在桌子上的纸牌,我们的任务是对它们进行分类。假如我们每次先用右手拿一张牌,然后将它放到左手的适当位置,这种做法便可以被看作一种用来解决这一任务的方法。这种方法实际上就是一种算法,在计算机科学里被称作插入分类法(insertion sort)。

值得引起我们注意的问题是,算法和方法或技术的概念之间存在的微妙的区别。例如使用的方法可能是最小平方法,其中使用的技术则可能是矩阵求逆,而下三角形矩阵-上三角形矩阵分解(LU decomposition)和斯特拉森算法(Strassen' s algorithm)则是两种可供我们选择的求矩阵的逆的算法。单独一种数据分析法可能会使用一种以上的算法。

不用算法就不可能写出统计软件,因此算法对于定量方法学者的重要性是确定无疑的。尽管界面友好的统计软件包使终端用户在用计算机解决大多数问题时已可不必自己动手构想算法,但假如他们能对算法问题有一定程度的了解,这将会对他们的计算机使用有一定的帮助。例如,在使用最大似然估计【Maxmum Likelihood Estimation】法时,我们可用一个初始估计值作为起点,某些初始值会遇到无收敛问题。有时在遇到这样的情况时,只要稍微对初始值作一些修改,就可以使无收敛问题迎刃而解。但这时如果研究者不具备算法的基本知识,则有可能被一个无收敛函数难倒。

图 1 显示了另外一种掌握一定的算法知识可能会对计算机使用有一定用处的情形。该图表示的是某一个网络数据的两种图示。左边的图版使用的是多维尺度分析

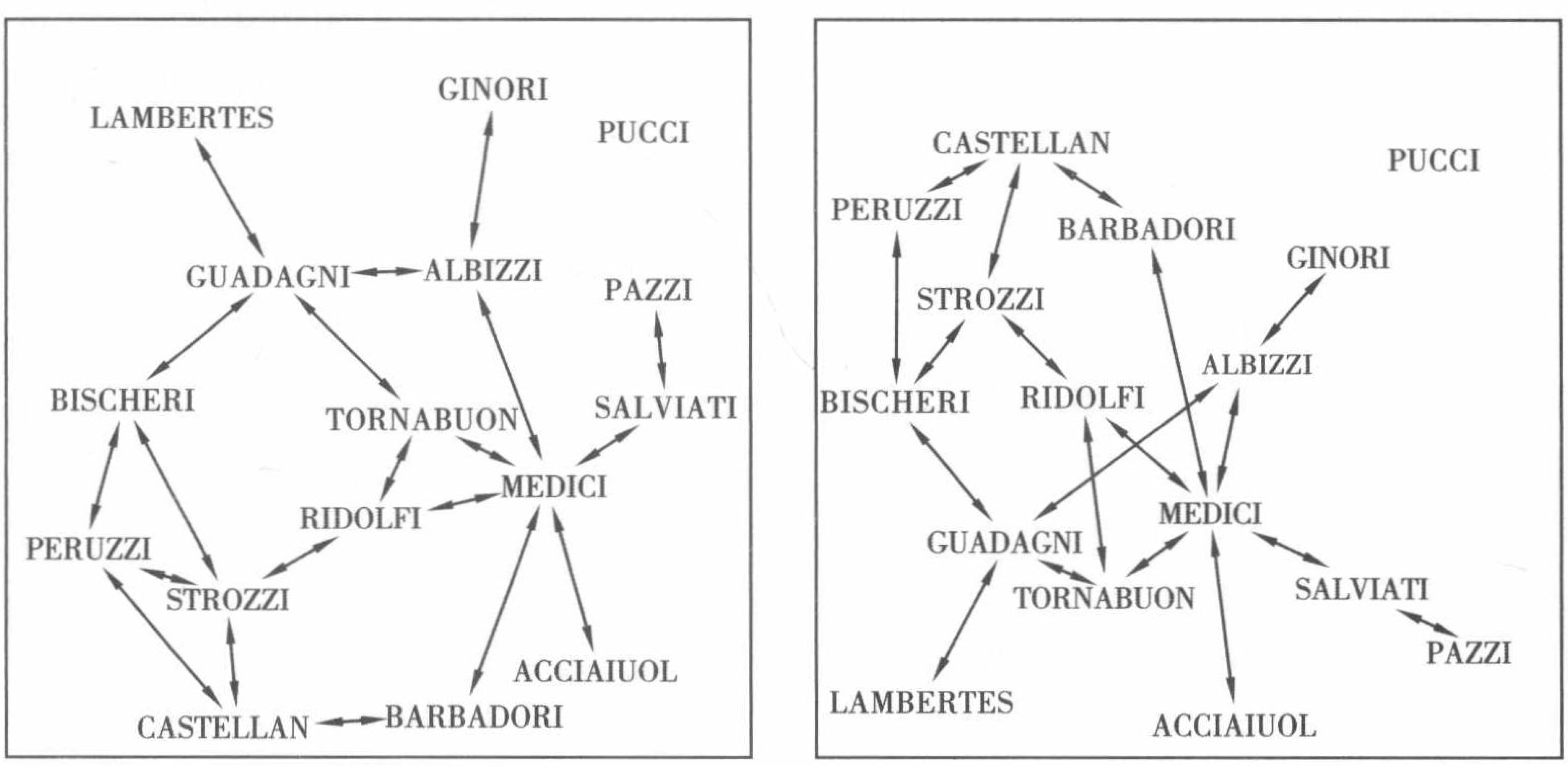



图 1　同一个网络数据的两种可能的图示

(multidimensional scaling),而右边的图版则使用了模拟退火(算法)(simulated annealing)。虽然两个图版使用的是同一个数据(它可以通过观察两图中人名之间的连接情况来加以验证),但它们的图形却不尽相同。在这里,算法是非常重要的,因为对网络数据的解释会受到图形的影响,而图形则会受到选用的算法的影响。虽然在许多时候,尽管不同的算法得到的结果都是相同的,但计算的速度却有很大差别。

算法一词的使用有时会更为宽泛,任何一种用一步一步的程序来解决某一给定的任务的方法,不论是否使用计算机都被叫作算法。例如,在对来自几个文档的历史记录进行匹配时,便可以手工使用一种算法。

——Andrew Noymer

(沈崇麟译校)

## 参考文献

Cormen, T.H., Leiserson, C.E., Rivest, R.L., & Stein, C. (2001). *Introduction to algorithms* (2nd ed.). Cambridge, MA: MIT Press.

Knuth, D.E. (1997). *Fundamental algorithms: The art of computer programming* (Vol. 1, 3rd ed.). Reading, MA: Addison-Wesley.

# Alpha($\alpha$),检验的显著性水平(Alpha, Significance Level of a Test)

显著性检验的发展要归功于罗纳德·费希尔。1920年代他从事土地统计工作,选择0.05来作显著性阈值(significance threshold),这一显著性水平后来为许多领域的研究团队所接受。研究者对假设检验和$p$值被人们接受的基本动机一直在进行着探讨(Goodman, 1999)。而我们必须强调的是,并没有任何数学理论表明,0.05的显著性水平的第一类误差水平(Type Ⅰ error level)是最优的,我们之所以使用这一阈值,仅仅是传统习惯所致。

阿尔法误差水平是一种在研究设计阶段选择的预期水平;而$p$值则是根据研究数据计算得到的误差水平。阿尔法水平之所以被选择,是因为它们指出了随机的抽样误差产生这一研究结果的可能性,因而它们可以被选择为0.05之外的其他水平。选择大于0.05作为阿尔法水平,意味着调查者愿意接受更大的第一类误差,以确保在自己的研究结果中有更大的得到正面效应的可能性。相反,如果在研究结果中发现的正效应可能非常重要,或可能引起这样那样的混乱,那么调查者就会设置一个更为严格的阿尔法误差率(例如$\alpha=0.01$)。如果这些决定是(a)有预期地作出的,(b)对学界作出了精确解释的,那么它们就被认为是恰当的。

一直以来始终在采用的一种显著性检验法是单尾检验法【One-Tailed Testing】。单尾检验法在医学研究领域一直备受争议,在社会科学界人们对它也褒贬不一。至于会导致非传统的阿尔法误差水平的双尾检验法【Two-Tailed Testing】,则在两个对称的显著性区域进行。在作改进学龄儿童教育的干预效应的检验时,有可能会出现自相矛盾的效应,干预降低而非提高了检验的分数。这时,调查者可以将第一类误差率拆分,于是得到的第一类误差水平的80%(如果总的阿尔法误差水平是0.05,那么它就是0.04)位于尾部的"有害"端,而将其余的1%则置于分布的有利的那一尾端。预告诸如这样的程序是至关重要的,因为它证明调查者对干预可能导致的不可预期的危害是敏感的。现在,用于发

展这种对称检验的指导方针也已经制订出来了(Moyé, 2000)。

在同一次临床试验中进行多个假设检验时,总的或整个的第一类误差水平便会上升。一项产生两个统计假设,得到在0.045水平的 $p$ 值的研究工作可能在告诉我们,研究样本产生了两个结果,它们很可能反映了那个样本从中抽取的较大总体的发现。而重要的问题在于,这些结论中,至少有一个是错误的。假定这两个检验的每一个都是独立的,那么这些分析中至少有一个会产生错误的结果,只由抽样误差本身造成的概率等于 $1-(0.955)^2=0.088$。这可以通过对单个检验的第一类误差水平进行调整来加以控制(即 Bonferroni 法【Bonferroni Techniques】)。这些调整对降低每个已经做过的统计假设检验的第一类误差水平是很有用的。举一个戒烟项目方面的例子,在这个项目中,调查者的兴趣在于考察戒烟项目在青少年烟民中的效应,他可以把 0.035 的第一类误差水平放在自我报告的抽烟方面的干预效应上,而把其余的 0.015 放在干预对于抽烟的态度的效应上。这种方法可以使总误差仍然保持在0.05的水平的同时,允许在抽烟量度两端的无论哪一端,都能使治疗(干预)效应有正向的结果。

——Lemuel A. Moyé

(沈崇麟译校)

## 参考文献


Goodman, S. N. (1999). Toward evidence-based medical statistics: 1. The *p* value fallacy. *Annals of Internal Medicine*, *130*, 995-1004.

Moyé, L. A. (2000). *Statistical reasoning in medicine*. New York: Springer-Verlag.


# 备择假设(Alternative Hypothesis)

一般我们都把与虚无假设【Null Hypothesis】对立的假设【Hypothesis】叫作备择假设。它通常都是研究者主导的假设。例如,我们的研究假设是食糖的食用与儿童多动症有关。那么假设 $H_A$ 就是虚无假设 $H_0$——食糖的食用与儿童多动症无关——的备择假设。

——Michael S. Lewis-Beck

(沈崇麟译校)

# 矩结构分析(Analysis of Moment Structures, AMOS)

AMOS 是一种在社会科学中用于结构方程模型【Structural Equation Models】或协方差结构模型【Covariance Structure Models】的统计软件。AMOS 有着良好的图形界面。有关它的更为详细的介绍请参考相关网站。

——Tim Futing Liao

(沈崇麟译校)

# 协方差分析(Analysis of Covariance, ANCOVA)

被称为协方差分析的方法源自一种将方差分析【Variance Analysis】和回归分析【Regression Analysis】合二为一的愿望。虽然在优秀的计算机统计软件问世之前,人们对它有很大的兴趣,但现在单独使用这种方法却越来越少。更为常见的是,这种分析只是简单地叫作多元回归分析,包含一个定量的因变量及作为解释变量的定量变量【Quantitative Variable】和虚拟变量【Dummy Variable】。

## 定义

协方差分析是一种多变量统计分析法,它的因变量是一个定量变量,而它的自变量却是混合型的,既有定类变量【Nominal Variable】,也有定量变量。

## 解释

方差分析用于研究一个定量的因变量和一个或多个定类的自变量之间的关系。而回归分析则用于研究一个定量的自变量和一个或多个定量的自变量之间的关系,这样马上就会出现一个问题:因为我们经常会遇到研究假设要求某些自变量是定类的,而另一些则是定量的这样的问题。随着统计计算机软件包的面世,这一问题很快便得到了解决。我们只需先将那些定类变量转化为一个或多个虚拟变量,然后再使用一个定序的多元回归分析就可以了。这样的分析法既可以使用,也可以不使用交互【Interation】变量。究竟是否需要使用,则取决于我们得到的数据和分析的问题的性质。现在人们越来越少单独使用协方差分析,因为回归分析可以轻而易举地处理它能处理的问题。

## 发展历程

1930 年代是协方差分析发展的重要时期,大量工作都是在那一时期完成的。因为罗纳德·费希尔爵士用方差分析作了大量的分析,而成为该方法发展过程中的领军人物。他的成就得到了威廉·科克伦(Cochran,1957)和其他六位在那一时期有重要影响的统计学家的赞赏。他们在《生物计量学》(*Biometrics*)杂志的一期特刊上,分别撰文推介罗纳德·费希尔在方差分析方面的出色工作。协方差分析是方差分析和回归分析这两种方法的自然延伸。在方差分析中,我们有理由控制一个定量变量,并对感兴趣的修正的而非原始均值进行比较。同样,在回归分析中,我们也有理由控制一个定类变量。正是这两种想法导致协方差分析和用台式计算器作简便计算方法的发展。

## 应用

最为简单的协方差分析的应用方式只用一个定量变量和一个作为解释变量的定类变量。用回归的记号,这一模型可写作

$$Y = \alpha + \beta_1 X + \beta_2 D \qquad (1)$$

式中的 $X$ 和 $Y$ 都是定量变量,$D$ 是一个含 0 和 1 两个组的虚拟变量,而 $\alpha$, $\beta_1$ 和 $\beta_2$ 则是回归参数。

这样,现在我们就可以作方差分析了,在控制 $X$ 变量的同时,比较两个组的均值。这时我们比较的是修正的 $Y$ 的均值,而非直接比较两个组的 $Y$ 的均值。通过将 0 和 1 代入虚拟变量,我们得到有以下方程的两条平行的直线:

$$Y_1 = \alpha + \beta_1 x_1 \quad (2)$$

$$Y_2 = (\alpha + \beta_2) + \beta_1 x_2 \quad (3)$$

$X$ 的一个公值，譬如说总均值，与 $Y$ 的预测值之间的差就变成了系数 $\beta_2$。于是，当我们将变量 $X$ 固定为一个常数时，修正的 $Y$ 均值的差，就变成了两条直线的垂直差异。

我们也可以通过控制定类变量 $D$ 对两个组进行比较。我们可以通过将数据分成两组，分别观察 $D=0$ 和 $D=1$ 作 $Y$ 对 $X$ 的回归。在两组有着完全相同的斜率时，我们就可以这样做，且我们已经设计出了在台式计算器上作这样的分析的详尽公式。但是较之分别进行两个回归，一个含两个自变量的多元回归则有着更高的效率。

这一方法可以推广至每个种类都使用多个解释性变量的场合。

## 例子

在一个小规模的学院，有人比较了男性和女性的收入。方法是比较男女两组的平均工资。比较结果发现，男性的均值大于女性。然而，人们还发现在教学人员中，更多的女性从事较低的教职，也许这才是导致男女收入差异的真正原因。为了控制年龄，我们可令年龄恒定，即先将数据按年龄分组，再在同一年龄组内比较男女的工资均值。由于我们着手分析的是一个小数据，这样我们就可能遇到有些分析只有很少几个个案这样的情况。

同样，我们也可以通过对每一性别分别进行分析来控制性别。这样分析可能就只会涉及两个组。尽管如此，这样的组仍然小于将整个教职员作为一个组的组。

相反，一个含年龄变量和虚拟的性别变量的多元回归就不会遇到小数据这样的问题。系数 $b_1$（$\beta_1$ 的估计值）将会显示控制了性别之后年龄的效应，而系数 $b_2$（$\beta_2$ 的估计值）则会显示在控制年龄之后性别的效应。

我们还可以作更为复杂的分析，允许收入在年龄上的回归线不是平行的。这种分析允许我们考察年龄和性别对工资的效应。要这样做，我们只要建立一个交互变量，它是年龄变量和虚拟变量的乘积。在将这一乘积变量引进多元回归方程时，该方程将工资作为因变量，而将年龄、虚拟和交互变量作为三个解释变量。

——Gudmund R. Iversen

（沈崇麟译校）

## 参考文献

Cochran, W. G. (1957). Analysis of covariance: Its nature and uses. *Biometrics*, *13*(3), 261-281.

Hardy, M.A. (1993). *Regression with dummy variables* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-093). Newbury Park, CA: Sage.

Wildt, A. R., & Ahtola, O. T. (1978). *Analysis of covariance* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-012). Beverly Hills, CA: Sage.

# 方差分析（Analysis of Variance, ANOVA）

在方差分析名下集合了一组用于一个或多个定类的自变量，对于一个定量的因变量的影响的统计分析方法。一个定类变量【Nominal Variable】的方差分析，叫作单因素方差分析【One-Way Analysis of Variance】，两个定类变量的方差分析，则称为双因素方差分析【Two-Way Analysis of Variance】，以此类推至更多因的模型。含有多个因变量的数据

可以用 MANOVA(多元方差分析【Multivariate Analysis of Variance】)。

## 解释

一项有关性别和种族对收入的影响的研究,采用方差分析来判别各种性别/种族组之间的平均收入是否有所不同。这些方法也许被称作均值分析(analysis of means)更为贴切,但是方差则是实际上用来确定组均值究竟是否不同,这种方法也因此而被叫作方差分析。

## 发展历程

方差分析发端于实验数据研究。许多早期的应用都在农业方面,研究的目的是弄清楚(诸如肥料和土地的类型)这些因素是怎么对一块土地上的作物产量产生影响的。英国统计学家罗纳德·费希尔在开发方差分析方法的早期做了大量工作。实验可以以许多不同的方式来设计,而从这些不同设计中得到的数据,则可用对应的方差分析法来分析。由此可见,在方差分析的名义下包含着数量众多的方法,而在这些方法中究竟应采取哪一种方法,则取决于实验设计采用的方式。实验设计和方差分析二者之间的关系是很强的。

方差分析和回归分析之间的关系也是很强的。从发展的历程看,这两套统计方法是互不相干的,工作在这两个领域的人们彼此联系很少。老的精确的方差分析计算方法,与回归【Regression】分析有很大的差别。但是现在,在计算方面,统计软件已经取代了手工或计算器,所以要证明这两套方法在许多事情上是相同的这一点并不是一件太困难的事。大多数方差分析都可以用含虚拟变量【Dummy Variable】的回归分析来表示。因此,我们的确可以把方差分析和回归分析看作我们称之为广义线性模型的特例。而这一点,在人们开始将方差分析法不仅用于实验数据,也用于观察数据时,就变得更加清楚。

## 应用和例子

### 单因素方差分析

方差分析的许多核心理念可用最简单的例子来阐述。这个例子有来自两个组的自变量 $Y$ 的数据。这就是说,我们有一个有两个值的定类的自变量。这样我们的问题就变成了一个单因素方差分析,因为数据只涉及一个自变量。这一数据可以来自某一心理学实验。这一实验有两个组,一个是控制组,另一个则是实验组。后者的成员将会受到某种刺激。而我们希望了解的问题则是,这些刺激是否有任何的效应。同样,数据也可以来自一个收集不同种族人口的收入的观察研究。我们想要了解的问题是,各个不同种族人口的平均收入是否存在差异。因此,我们知道一个成分(人)属于什么组,而对于每一个成分,我们则有其在定量变量 $Y$(响应,收入)上的数据,为了便于在电子计算机软件开始使用后数字计算的方便,我们使用方差分析中普遍使用的记号,它看起来与用于回归分析的记号有很大的不同。

一般我们用 $y_{ij}$ 来表示变量 $Y$ 的观察值。第一个下标 $i$ 表示观察所属的组,在这里 $i=1$ 或 $i=2$。在每一个组内,观察值都作了编号,编号则用 $j$ 来表示。例如,$y_{25}$ 则表示第二组中的第 5 个元素的观测值。同样,第 $i$ 组的均值则表示为 $\bar{y}_i$,而所有观察的均值则表示为 $\bar{y}$。用这些符号,我们便可得到如下简单的恒等式:

$$y_{ij} - \bar{y} = (\bar{y}_i - \bar{y}) + (y_{ij} - \bar{y}_i) \quad (1)$$

观察值 $y_{ij}$ 与总均值 $\bar{y}$ 之间的差,则被表示为组均值 $\bar{y}_i$ 与总均值 $\bar{y}$ 的差加观察值与组均值的差之和。

如果人们的收入不受一个人所在的组，或任何其他变量的影响，那么所有的观察值都应该与总均值相等。每个人都应该有相同的收入。我们就绝不可能得到不同的观察值，因为它们不受任何变量的影响。那么观察到的围绕总均值 $\bar{y}$ 的观察值的变化可能是由两个因素所致：分类的自变量和所有其他的变量。

同样，上式等号右边的第一项测量的便是自变量的影响有多大。如果这个变量没有影响，那么两个组均值便应该相等，同时如果所有其他变量都没有影响，那么每组的观察就应该全部相等，组均值和总均值之间的差便等于零。等号右边的第二项测量的是每一个观察值与组均值的差异。如果所有其他的变量都没有影响，那么每一组中的观察值就应该全部相等，因为只有组的成员资格才是唯一有关的因素。因而观察值和组均值之间的差的确是其他变量，即所谓残余变量的影响。在总体中，我们则把它叫作误差变量，而它的样本估计值则形成了残余变量：

$$\frac{\text{自变量和残余}}{\text{变量的联合效应}} = \frac{\text{自变量}}{\text{的效应}} + \frac{\text{残余}}{\text{效应}} \qquad (2)$$

这一恒等式适用于单独一个观察值。下面我们想来概括一下跨所有观察值的效应。如果我们把所有的效应加起来，那么这三项的和都等于零，因为这些效应是对均值的全部偏差。

效应的量

而要得到正的项，我们则需取上式的左右两边的平方。用平方无非是历史的和数学上的方便。使用平方数实际上没有对任何东西做"纠正"，但必须认识到我们分析得到的所有结果将也会因为选择平方而受到很大影响。

将左侧的平方相加得到联合效应的平方和。将右侧的平方和相加，则得到自变量效应的平方和，加上残余变量效应的平方和，再加上自变量和残余变量效应的积的和乘以2。由此可见交叉乘积的和总是等于零。

取平方后我们得到如下的跨所有观察值的平方效应方程，用符号、文字和缩写我们可将此方程写作：

$$\sum\sum(y_{ij}-\bar{y})^2 = \sum\sum(\bar{y}_i-\bar{y})^2 + \sum\sum(y_{ij}-\bar{y}_i)^2$$

$$\text{总平方和} = \frac{\text{组间}}{\text{平方和}} + \frac{\text{组内}}{\text{平方和}} \qquad (3)$$

$$TSS = BSS + WSS \qquad (4)$$

平方前面的双求和符号表示我们先将每一组内的所有的观察项相加，然后再将两组的项相加。

可见对每一个体而言，我们将联合效应拆分成自变量和残差效应的平方和。借助取平方，我们也可以将个体的整个组的联合效应拆分成自变量效应（组间平方和【Between Sum of Squares】）和残差变量效应。对于更为复杂的方差分析而言，我们将进行更大量的总平方和【Sum of Squares】的拆分。

为了简便起见，我们可以将总平方和分成三个平方和，这将给我们如下所示的代数式：

$$\frac{TSS}{TSS} = \frac{BSS}{TSS} + \frac{WSS}{TSS} \qquad (5)$$

$$1.00 = R^2 + (1 - R^2) \qquad (6)$$

这一步确定了决定系数 $R^2$【Coefficient of Determination $R^2$】。它告诉我们自变量效应在总平方和中的比例。在以往老式的记号法中，这一个量常用 $\eta^2$（eta$^2$）来表示。它与回归分析中的 $R^2$ 起的作用是一样的。$R$ 本身是一个系数，用以测量自变量和因变量之间的关系强度。上述数量经常会在方差分析表中展示，实际就如表 1 所示。

表 1 平方和与比例

| 效应源 | 平方和 | 效应比例 |
| --- | --- | --- |
| 自变量 | $\sum\sum(\bar{y}_i-\bar{y})^2$ | $R^2$ |
| 残余变量（误差） | $\sum\sum(y_{ij}-\bar{y}_i)^2$ | $1-R^2$ |
| 总和 | $\sum\sum(y_{ij}-\bar{y})^2$ | 1 |

## 假设检验

这种关系是否偶然发生？这一问题可转换成一个有关总体均值的虚无假设：

$$H_0: \mu_1=\mu_2 \quad (7)$$

在含有两个以上的总体时，问题将会变得更加复杂。这个虚无假设的陈述是，所有总体的均值都相等，而与之逻辑相反的陈述则只能是，至少有些或并非所有的均值都相等。一旦虚无假设被否定，问题马上就变成了究竟哪些是不等的。

研究虚无假设的方法是求出与之对应的 $p$ 值【$p$ Value】。诸如这样的 $p$ 值告诉我们，得到来自总体的这样的数据，或某些更为极端的数据，证明虚无假设可能为真的可能性有多大（次数的比例）。这可以被看作否定虚无假设的概率。如果 $p$ 值比较小——譬如说小于 0.05——我们就拒绝虚无假设。求数据 $p$ 值的方法是计算相应的统计学的 $F$ 变量的值。而实际的计算过程，只有统计学家才对此感兴趣，故而不再详述。表 2 中列出与本表相对应的 $F$ 值，它几乎可用大多数统计软件计算得到。

表 2 使用平方和、比例、均方和 $F$ 值的方差分析

| 效应源 | 平方和 | 效应比例 | 自由度 | 均 方 | $F$ 值 |
| --- | --- | --- | --- | --- | --- |
| 自变量 | $\sum\sum(\bar{y}_i-\bar{y})^2(BSS)$ | $R^2$ | $k-1$ | $\frac{BSS}{k-1}=BMS$ | $\frac{BMS}{WMS}$ |
| 残余变量 | $\sum\sum(y_{ij}-\bar{y}_i)^2(WSS)$ | $1-R^2$ | $n-k$ | $\frac{WSS}{n-k}=WMS$ | |
| 总和 | $\sum\sum(y_{ij}-\bar{y})^2(TSS)$ | 1 | $n-1$ | | |

表 2 的列数多于表 1，它可以用来列 $p$ 值。字母 $k$ 表示组数，而字母 $n$ 则表示研究涉及的观察总数。

$F$ 的计算基于组间和组内平方和之比。但是平方和不能直接进行比较，因为它们取决于和数中的项数的多少。而将和除以它的适当的自由度【Degress of Freedom】，则每个平方和都转变成一个方差。这一转换得到了列表中为均方（mean square）的那个量。不过，该列却并不能再用方差这一词来作为列标。顾名思义，均方是“平均的”平方，而方差是平均的背离均值的平方。最后，这两个均方之比则以 $F$ 来表示（以统计学家费希尔的名字命名）。

通过上面的数学推导，我们可以证明，在虚无假设为真时，这两个均方（方差）大致相等。因此，这两个均方之比（$F$ 值）近似于 1.00。在虚无假设为假时，和总体均值的确不同，$F$ 值将远大于 1.00。

在残余变量的值服从正态分布时，$F$ 变量便会有相应的 $F$ 分布【$F$ Distribution】（一个类似于正态分布【Normal Distribution】$t$ 和卡方分布【Chi-Square Distribution】变量的变量）。然后我们可以用 $F$ 分布来求数据的 $p$ 值。$F$ 的临界值取决于其中有多少组和多少观察。而最终求解结果则要看 $F$ 的两个自由度，它等于$k-1$和 $n-k$。假如我们有一套来自两个组，总观察值数为 100 的和 $F=$

5.67的数据。而从 $F$ 分布表中我们发现,与自由度为 $2-1=1$ 和 $100-2=98$ 的 $F$ 值对应的 $p=0.02$。因为它比较小,我们否定了虚无假设,于是我们得出两个组的均值的差异具有统计显著性的结论。大多数统计软件都可以为我们提供确切的 $p$ 值。一般的统计教科书都附有 $F$ 变量表,但是因为每一对自由度的 $F$ 分布都不同,所以尽管 $F$ 表已经变得很大,但却仍然不够详细。

来自两个组的数据也可以用 $t$ 检验,或一个带有虚拟变量的回归分析,来分析两个均值之间的差异。这三种分析方法(双侧)的 $p$ 值都相等。如果我们有两个以上的组,那么就没有什么 $t$ 检验法可以对它进行分析,但是我们仍然可以用含适当数目的虚拟变量的回归分析,而不是方差分析。

### 双因素方差分析

在这里,两个定类自变量在影响一个定量的因变量。同样也可以有三因或更多因的分析,究竟有多少因则取决于有多少个自变量。

除了有一个以上的自变量之外,现在我们还有一个存在于自变量的(主)效应之外的所谓的交互效应。自变量的交互效应是一种基于两个单独的自变量,但超越两个单独的自变量效应的效应。

分析基于下面的恒等式:

$$y_{ijk}-\bar{y}=(\bar{y}_{i.}-\bar{y})+(\bar{y}_{.j}-\bar{y})+(\bar{y}_{ij}-\bar{y}_{i.}-\bar{y}_{.j}+\bar{y})+(\bar{y}_{ijk}-\bar{y}_{ij}) \tag{8}$$

在这种记号中,每一观察值都用三个下标 $i$、$j$ 和 $k$ 表示。第一个下标给出第一个定类变量的值,第二个下标给出了第二个定类变量的值,而最后那个下标则给出了两个定类变量的特定组合的观察数。

如果没有东西对因变量有影响,那么该变量的所有的值将都等于总均值 $\bar{y}$。如果性别、种族和其他变量都没有受影响,那么每一个人就会有相同的收入。与这一个值的任何偏差都代表一或若干个自变量是有效应的,这就如上式等号左侧所示。等号的右侧,第一个差表示第一个定类变量的效应,第二个差则表示第二个定类变量的效应,第三项表示交互变量的效应,而最后一项则表示残差,它测量的是所有其他变量的效应。如果不包括交互项,那么等式就无法成立。

对上面的方程式取平方,右侧给出了每一项的平方加上 2 倍的交叉乘积。将所有的观察值加在一起,所有交叉乘积的和便等于零,得到的平方和则给出了自变量和残余变量的效应。这些和数可以比例来表示,而将它们除以残差的均方,我们便可以得到 $F$ 值,从而得到无效应假设检验的 $p$ 值。

### 模型Ⅰ、模型Ⅱ和模型Ⅲ:固定或随机效应模型

在我们有所有的自变量值的因变量数据时,我们便会有模型Ⅰ,ANOVA(固定效应)。在我们只有自变量的样本值的因变量数据时,我们则有模型Ⅱ, ANOVA(随机效应)。模型Ⅲ,ANOVA 则是固定效应和随机效应的混合体。

——Gudmund R. Iversen

(沈崇麟译校)

## 参考文献

Bray, J. H., & Maxwell, S. E. (1985). *Multivariate analysis of variance* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-054). Beverly Hills, CA: Sage.

Cobb, G. W. (1998). *Introduction to design and analysis of experiments*. New York: Springer-Verlag.

Iversen, G. R., & Norpoth, H. (1987). *Analysis of variance* (2nd ed., Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-001). Beverly Hills, CA: Sage.

## 分析式归纳(Analytic Induction)

分析式归纳有时被作为定性研究设计【Research Design】的一种模型对待。它把调查看作由以下环节组成的迭代过程:对要解释的现象作初始定义;考察若干个案;建立假设【Hypothesis】;对个案作进一步的研究以对这一假设进行检验;假如证据否定了假设,那么或者重新对现象作定义,以排除否定的假设,或者重新建立假设以能对它作出解释。接着,我们再对那些个案作进一步研究,如果这进一步的证据支持假设,那么调查就可以得出结论。但如果将来出现了任何否定的个案,那么调查仍需重新开始,或者必须重新建立假设,或重新定义现象,然后对个案作进一步研究,如此等等。

归纳式分析与扎根理论【Grounded Theory】有共同之处,而与来自纸上谈兵的理论归纳的假设检验(参见归纳【Induction】)则不同。它也与一种强调通过对为数不多的个案,并采用一种灵活的使理论【Theory】得以显现的操作模式——并由此对下一步的调查,包括数据收集和分析作出指引的方法相通。此外,它也与许多自我理解式的定性研究【Qualitative Research】理论的特性相左,尤其是在注重假设检验和致力于发现这一点上,差别更为明显。它的最为显著的特点是,它认为被解释的现象类型可能需要在研究过程中重新予以阐述。这一观点的提出是因为它认为,调查的目的在于发现理论上的类别。所有理论上类别范畴的实例,都能用单独一类的原因作解释。在分析式归纳的语境中,一个理论同时确定了必要和充分两个条件,正因为如此,只要有一个例外,就足以推翻任何理论。

分析式归纳一词似乎是由波裔美国社会学家弗洛里安·兹纳涅茨基在他的著作《社会学方法》(*The Method of Sociology*, Znaniecki, 1934)一书中提出的。在书中,他将它与枚举归纳(enumerative induction)一词作了对比。枚举归纳一词涉及某种推论,是统计方法的一个特性。枚举归纳试图发现某类现象属于特定的类别。做法是先对大量的个案进行研究,进而对它们的共性加以描述。在这里弗洛里安·兹纳涅茨基指出,这一方法中存在着一个悖论:如果我们不了解某类现象的基本性质,我们就无法确认它的实例;然而如果我们确实已经知道这些基本性质,那么我们就已经知道枚举归纳要告诉我们的东西。有鉴于此,弗洛里安·兹纳涅茨基进一步论证,与当代某些社会科学家的认识相反,统计方法并非一种自然科学方法。自然科学并不研究大量的个案,并从它们的平均特征导出理论。相反,他建议要深入地研究单独一个,或少数几个个案,进而发展出将这些个案确认为一般的现象类别实例的基本特性的理论。故而,根据弗洛里安·兹纳涅茨基的观点,科学的任务是寻找普遍的规律,即因果关系或函数依赖,而非统计的概括。而这就是分析,而非枚举归纳方法,它的方法应当是归纳(Gomm, Hammersley, & Foster, 2000; Hammersley, 1989)。

兹纳涅茨基没有提供任何具体分析式归纳的例子,但是其后的两位作者却提供了现在认为是对分析式归纳的关键性的例解。艾尔弗雷德·林德史密斯(Lindesmith, 1968)对人们之所以会对鸦片上瘾的因果关系进行了调查之后,证实这些人变得有意识地服用毒品的基本原因是减轻断瘾症状(withdrawal symptoms)。他的一个学生,唐

纳德·克雷西则提出了确认信任破坏(trust violation)的必要和充分条件,这一概念导致侵占、挪用和贪污等问题的常识和法律概念的理论重建。而重建就发生在他的研究进行的过程中(Cressey, 1950, 1953)。他提出了一套相当复杂的理论,依据这些理论,当那些身居金融信托公司要职的人遇到不可对他人言的金融问题,并认为这个问题可通过侵占、挪用和贪污来秘密解决;即便他们可能求助于陈情、道歉,仍然可以把他们看成不值得信任的。

学界对分析式归纳的特征和价值一直争论不休。其中早期的但至今仍有价值的批评,是由鲁宾逊(Robinson,1951)提出的,他对枚举和分析式归纳二者的区别提出质疑,他认为分析式归纳无非就是一般的实际操作,它的全部意义几乎就是发现必要而非充分条件。在他看来,这是一个那些旨在以某种理论预测将要发生的现象的特定类型的研究者所无法克服的弱点。与此同时他还认为,要求展示普遍而非概率分布的证据,并要求所有的实例都与理论一致,没有任何例外,这样的标准未免苛刻,因为如果研究的个案数目过大,那么这样的要求几乎是无法满足的。拉尔夫·图尔纳(Turner,1951)则对这一批评持反对意见,他也强调分析和枚举归纳之间存在差别,但他认为构建解释框架时,二者可相互补充。虽然分析式归纳这一概念现在已经鲜有定量研究者问津,但霍华德·贝克尔(Becker, 1998)最近却反复提到这一概念,并对这一概念有所发展。他认为分析式归纳这一概念切中社会科学调查的要害。

——Martin Hammersley

(沈崇麟译校)

## 参考文献

Becker, H. S. (1998). *Tricks of the trade*. Chicago: University of Chicago Press.

Cressey, D. (1950). The criminal violation of financial trust. *American Sociological Review*, *15*, 738-743.

Cressey, D. (1953). *Other people's money*. Glencoe, IL: Free Press.

Gomm, R., Hammersley, M., & Foster, P. (Eds.). (2000). *Case study method*. London: Sage. [Includes the articles by Robinson and Turner.]

Hammersley, M. (1989). *The dilemma of qualitative method*. London: Routledge Kegan Paul.

Lindesmith, A. (1968). *Addiction and opiates*. Chicago: Aldine.

Robinson, W. (1951). The logical structure of analytic induction. *American Sociological Review*, *16*(6), 812-818.

Turner, R. H. (1951). The quest for universals. *American Sociological Review*, *18*(6), 604-611.

Znaniecki, F. (1934). *The method of sociology*. New York: Farrar and Rinehart.

# 匿名(Anonymity)

匿名是对研究参与者的一种保证,保证通过伪装或隐去他们的个人特征,使其身份不会被辨识出来。在开展社会科学研究过程中,保护“人类被试者”这一问题,是机构审查委员会和研究人员等都十分关切的问题。而保证和保护个人的匿名性是保护他们作为研究参与者的权利的关键要素。

作为研究者,我们会采取若干措施来确保参与者的匿名性。例如,在识别他们的资料时,我们为每个人分配一个案例编号来代替被调查者的真实姓名。在呈现我们的研

究结果时，我们会隐藏所有可能使被调查者身份暴露的信息。例如，我们使用假名来代替真实姓名。研究者在对整套资料进行建构时，必须在删除所有个人的身份标识（如姓名、地址、社会保险号）之后，才能允许资料公开使用。

当然，我们通常都会注意对个人的直接身份标识进行处理，但间接的身份标识，例如居住地、组织的会员资格和职业，也能用来辨认某个人或群体（Inter-University Consortium for Political and Social Research, 2002）。定量研究者会对变量进行例行的再编码，将它们纳入更宽的类别，因为更宽的类别会恰当地捕捉到一些共享特征，而他们相信这些特征会影响他们感兴趣的结果变量。例如，关于内科医生的样本资料，可能包含他们获得医学学位的医学院类型（如私立或公立机构），而不是这些医学院的具体名称。在定性研究中，间接的身份标识可能带来更大的问题，部分原因在于定性资料不进行资料的合并处理。从事定性研究的社会科学家都力求针对他们所研究的群体，提供丰富的描述性资料。这种描述的丰富性在一定程度上可以使作者将个人的社会环境，与他的经历和认知联系起来。因此，我们通常希望读者了解一个特定被调查者的职业，以及职业可能以怎样的方式影响他或她生活的其他方面（或其他感兴趣的变量）。

保证研究参与者的匿名性，对于处理那些对科学研究并不信任的群体，以及那些曾参与过非法活动，或被社会污名化的活动的群体来说是尤为重要的，这类活动的一个例子是对药物的非法使用。那些曾经受到既往研究项目侵害或利用的群体，很可能就会对研究产生不信任感。历史上，为了研究梅毒的危害而拒绝对非洲裔美国人提供治疗的塔斯基吉梅毒实验（The Tuskegee Syphilis Study），就是这类研究中一个恶名昭著的例子。

在我对感染艾滋病病毒（HIV/AIDS）的女性进行研究的过程中，发现了许多让这些女性对保护她们的匿名性有所担忧的原因，比如非法的药物使用，她们子女的犯罪行为，以及将她们的艾滋病诊断结果透露给她们的家人和朋友，等等。有趣的是，另一方面，有些研究参与者并不希望保持匿名，因为她们是自豪的——为自己所克服的困难自豪，为自己在经历这个过程时的收获和成长而自豪，为自己是什么样子或者变成什么样子而自豪——因而她们愿意分享自己的故事。格里耶（Grinyer, 2002）曾报告过，一个参与她研究的癌症病人希望使用自己的真名而不是假名。尽管一些参与者有这样的意愿，但总的来说研究者还是有义务采取所有可能的措施来确保研究资料保密，并为研究参与者提供匿名处理。

——Desirée Ciambrone

（郑晓娟译　沈崇麟校）

## 参考文献

Grinyer, A. (2002). The anonymity of research participants: Assumptions, ethics, and practicalities. *Social Research Update*, *36*, 1-4.

Inter-University Consortium for Political and Social Research. (2002). *Guide to social science data preparation and archiving* [online]. Supported by the Robert Wood Johnson Foundation. Available: www.ICPSR.umich.edu.

## 应用定性研究(Applied Qualitative Research)

应用定性研究通常指那些显然是用于政策制订的定性研究,常见于政府部门,但也在商业中使用。这个术语可能于1985年在英国首创,是鼓励政府使用定性研究的尝试之一(Walker, 1985)。在那个时代,尽管定性方法论已经成功地进入了市场营销和广告研究领域,但政府官员仍倾向于将证据等同于官方统计【Official Statistics】数据(Weiss, 1977)。应用定性研究通常由政府机构或企业委托,研究的开展一般遵循合同规定的具体内容。

应用定性研究是一项具体活动,而不是一个运动或一套连贯的理念(Ritchie & Lewis, 2003)。这种研究的目标常常是用定性技术作为其他研究方式的替代或补充,来开发、监控或评估一项政策或实践。使用定性方法的理由在于定性方法在某些情境下,或在本质上更适合某项研究任务。这可能是因为该研究主题没有被清晰界定,或没有被很好地进行理论化,因而排除了使用结构化问卷进行研究的可能。研究的问题可能是敏感的、难以理解的、短暂的,或者可能与让人痛苦的、动情的事件,以及深层次的情绪有关。研究的问题可能本身十分复杂,或已经成为一种习俗惯例。针对这类问题,前期研究可能得出了相互矛盾的结论。

应用定性研究的实践者会用不同的研究方法匹配不同的研究问题,因而他们倾向于将自己与实用主义联系起来。他们的本体论定位可能与“微妙的实在论”相符(Hammersley, 1992),他们认为世界是多样性的和多维度的,它独立于人们的主观理解而存在,但能够通过被访者的解释来接近和到达这个领域。他们力求在研究过程的每个阶段做到中立和客观,继而得出有效可靠的结论,这样的结论能够既符合被访者自身的理念和理解,也适用于其他更广泛的场景。他们会探寻被访者所理解的事实的细节描述,同时在分析中清晰呈现这些解释与事实本身的差异,通过简化和结构化使复杂的事物状态言之成理。

应用定性研究在使用具体的研究方法方面是非常兼容并包的,但最为倚重的则是深度访谈【In-Depth Interview】和焦点小组【Focus Group】。目前有一种趋势是对同样的被访者进行重复的访谈,以此发现意图、行为与结果之间的关联。研究人员经常使用投射技术【Projective Techniques】,其目的并非得出在人格特质方面的结果分数,而是刺激被研究人员在参与研究过程中的反应和对投射行为的反馈结果;研究人员有时也会使用民族志【Ethnography】或行动研究【Action Research】的技术。在同一个项目中,不同的研究方法,包括定量研究方法,都会被研究人员使用。在进行定量调查过程中,研究人员使用定性技术可以达成多种目标,包括界定概念、设计问卷;对调查结果进行辅助的解释、阐述、说明和确证;包括在不同研究方法之间获取平衡以及提供互补性的洞见。

——Robert Walker

(郑晓娟译　沈崇麟校)

### 参考文献


Hammersley, M. (1992). *What's wrong with ethnography*. London: Routledge.

Ritchie, J., & Lewis, J. (2003). *Qualitative research practice*. London: Sage.

Walker, R. (Ed.). (1985). *Applied qualitative research*. Aldershot, UK: Gower.

Weiss, C. (1977). *Use of social research in public policy*. Lexington, MA: D. C. Heath.

# 应用性研究(Applied Research)

应用性研究从定义看,指的是那些基于实用性理由而开展的,因而通常会很快得到应用的研究。研究结果通常是可操作的,并会在特定领域带来提升效果和效率的有利变化。应用性研究很少致力于推进理论体系的发展。在应用性研究中,理论是一种用来确立和界定概念及变量的工具,以便研究者可以为了研究而对概念和变量进行操作化。

应用性研究与基础研究【Basic Research】(有时也称为纯研究或纯学理研究)的概念是相对的,后者在研究完成后并不能带来迅速直接的应用。基础研究的目的并不在于推动实际的或快速的应用,而在于推动我们对基本的社会过程,以及社会互动的理解和认识。从长期看,几乎所有的基础研究最终都会产生有价值的应用。二者的主要区别在于,研究是否可以得到迅速直接的应用。不过,应用性研究更容易与大众相关,因为它产生可付诸行动的和实用性的结果,而基础研究直觉上给人的感觉是只与研究者有关。

应用性研究的一种常见类型是项目评估。在某些案例中,研究者研究现行项目的有效性,以便评价该项目在达成目标方面有多大效果。在另外一些案例中,研究者可能研究项目推进系统的效率,以检验资源浪费的程度。在无论哪一种案例中,研究者的目的都是发现改善项目或其推进系统某些方面的方法。

应用性研究的另外一个例子是市场研究。市场研究人员可以通过研究(焦点小组、深度访谈或问卷调查)来探寻消费者的产品需求和偏好,进而包装出最合适的产品。市场研究人员也可以在消费者中测试他们的营销宣传材料,以确保他们想表达的营销信息能够有效地传达。这种类型的研究所产生的建议包括改变用词、包装风格、物流路径,甚至是宣传材料的颜色使用方案。

在许多案例中,应用性研究的受众通常局限于特定的利益相关方。因此,应用性研究的结果通常作为发行量有限,或独家所有的内部报告来处理。但是,有些应用性研究会通过在专业期刊上的发表来实现研究成果的共享。例如,《老年学刊》(*The Gerontologist*)除了刊登常规长度的文章外,还有一个名为"实用概念"的专栏,专门呈现应用性研究和一些篇幅较短的报告。"实用概念"曾刊登了一篇由西、诺伊费尔德和邓巴(Si, Neufeld, & Dunbar, 1999)三位作者所写的文章,研究在短期的护理康复中心,安装床栏与居住者安全之间的关系。西等人发现,在一些选定的护理康复中,大量减少床栏的使用并不会导致居住者受伤情况的增加。他们认为,床栏的必要性取决于护工人员接受的训练和经验,同时与病患受伤的类型和严重程度有关。

如果想对应用性研究作进一步的了解,请参考以下书籍:亨德里克等1993年出版的《应用性研究设计:实践指南》(*Applied Research Design: A Practical Guide*);克鲁格2000年出版的《焦点小组:应用性研究实践指南》(*Focus Groups: A Practical Guide for Applied Research*)。

——VoonChin Phua

(郑晓娟译　沈崇麟校)

## 参考文献

Hendrick, T., Bickman, L., & Rog, D. J. (1993). *Applied research design: A practical guide*. Newbury Park, CA: Sage.

Krueger, R. A. (2000). *Focus groups: A practical guide for applied research.* Thousand Oaks, CA: Sage.

Si, M., Neufeld, R. R., & Dunbar, J. (1999). Removal of bedrails on a short-term nursing home rehabilitation unit. *The Gerontologist*, *39*, 611-614.

## 档案研究(Archival Research)

对于社会科学家来说,档案研究可以被定义为对档案中发现的资源的定位、评价、系统阐释和分析。原始来源的材料是可以参考和分析的,但分析目的则不同于它们在最初搜集的时候。分析原始来源的材料的目的是对旧数据提出新问题;提供跨时代或跨地域的比较;证明或挑战已有的发现;或是从不同的来源中聚集证据,以描绘更恢宏的图画。

对社会科学家而言,使用档案似乎没有进行田野调查那么令人兴奋。在很多研究者眼中,田野调查新鲜、生动,极为重要,而历史研究的方法和档案技术似乎不那么常见,例如,社会科学家可能不习惯在故纸堆上工作,而这对历史学家来说则是家常便饭。不过,这些方法应该被当作是互为补充的,而非相互竞争的,它们在处理某些研究问题时具有的优点应得到承认。实际上,查考档案材料能使社会科学家得以改进和挑战用以界定和搜集数据的既有方法。

那么,社会科学家对哪种材料感兴趣呢?档案包括官方材料(例如政府文件)、组织记录、医疗记录、个人收藏以及一些情境性的材料。基础的研究数据也包括在档案中,例如调查者在研究过程中创建的数据,包括访谈的录音整理稿或录音带、田野笔记【Fieldnotes】、个人日记【Diaries】、观察报告、未刊稿以及一些相关的通信。这类材料在许多大学资料库或私人收藏中都极为丰富。档案还包括在制度、理论或思想变迁进程中的文化或物质遗存,例如在一个重要的社会科学部门里的思想发展。

不管怎样,档案肯定是多年沉淀的产物——收藏品可能会遭到侵蚀或破坏——可能是自然的(例如意外丢失或损毁),也可能是人为的(例如有意的分割或机构的处理政策)。因此,材料是否有保留价值取决于收藏者或档案保管员的主观判断,因此可能不能代表原来收藏的全貌。对档案馆来说,储藏空间也会对其起初决定收藏或拒绝什么藏品有很大的影响。

搜寻档案材料的技术类似于挖洞——你能分析什么取决于你能找到什么。一旦搜寻到了材料,研究者就可以自由地把自己浸泡在材料中——去评估、审查和重新分类数据;去检验之前的假设;或者去发现浮现的议题。评估原始材料的信度和效度是准备和分析材料至关重要的第一步。数据材料筛选、集中起来之后,就可以运用许多传统的社会科学分析方法——例如使用扎根理论【Grounded Theory】来探求数据的模式和主题;通过核查个人文献材料,运用传记法【Biographical Method】来记录个人生活和思想;或是用内容分析【Content Analysis】来确定文本中某些字词或概念的存在。

获准查看档案不应当被看作是理所当然的,为了查考档案材料,研究者一般必须先行预约,并出示身份证明。对有些收藏而言,还必须事先取得档案保管员的特别许可。

——Louise Corti

(杨可译　沈崇麟校)

## 参考文献

Elder, G., Pavalko, E. K., & Clipp, E. C. (1993). *Working with archival lives: Studying lives* (Sage University Papers on Quantitative Applications in the Social Sciences, 07-088). Newbury Park, CA: Sage.

Foster, J., & Sheppard, J. (1995). *British archives: A guide to archival resources in the United Kingdom* (3rd ed.). London: Macmillan.

Hill, M. M. (1993). *Archival strategies and techniques* (Qualitative Research Methods Series No. 31). Newbury Park, CA: Sage.

Scott, J. (1990). *A matter of record: Documentary sources in social research.* Cambridge, UK: Polity.

## 定性数据存档(Archiving Qualitative Data)

定性数据存档是以能够被现在和将来的研究者取用的格式长期保存的定性数据。保证数据长期可取用性的关键在于数据处理、创设信息文件的策略,以及储存、保护、安全保障和存取的物理和技术规程(参见数据归档【Data Archives】)。

定性数据【Qualitative Data】是采用定性研究【Qualitative Research】的方法和各种跨社会科学学科的技术搜集的资料。定性研究通常都包括多样化的,而非单一的方法和工具,其搜集的数据类型取决于研究目的、样本性质和具体的学科。因此,数据类型包括深度访谈【In-Depth Interview】或非结构访谈【Unstructured Interview】、半结构化访谈【Semistructured Interview】、田野笔记【Fieldnotes】、无结构日记【Diaries】、观察报告、个人文献和照片。定性研究经常会生成大量的原始数据,尽管这些研究方法的样本量一般都比较小。最后,这些数据可能是以一些不同的格式建立的:数字化的、纸质的(打印或手写的)或是音像材料。

1994 年之前,尽管口述史研究团体对保存口述历史访谈项目中收集的录音带已经有专业兴趣,但对定性数据却一直还没有系统的存档和传播规程。许多这类录音带可以在伦敦的英国国家图书馆音响资料馆(the British Library National Sound Archive in London)找到。

1994 年,在经济与社会研究理事会(Economic and Social Research Council)的支持下,英国首先设立了国家级的定性数据存档项目。定性数据中心(Qualidata Centre)为原始数据和伴生文献(元数据,meta-data)的分类、描述、处理建立了常规程序,并为数据的存取使用建立了适当机制。

处理数据集的方法可以分成三个主要工作:核查与匿名处理、转换数据集和生成元数据。核查数据集包括确认数据的完整性和质量,以及在数据项(如访谈、田野笔记和录音)之间建立关联。优质的录音整理稿对于将没有录音的访谈收集归档至关重要,同样,保证数据内容符合与参与者事先达成的所有许可协议这一问题也很重要。保密【Confidentiality】和版权是定性数据在存档和传播中出现的两个主要问题。数据转换指的是将数据转成适于保存和传播的格式,这可能涉及对纸质材料的数字化处理。

第三项主要的处理工作是生成元数据(关于数据的数据),指的是处理中生成的背景信息,比如,生成一个包含访谈细节的"数据表",确保在原始文本上加上讲话者、访问者的姓名标签,以及问题或主题指南标题。元数据的关键功能之一,就是让

使用者能在一个数据集中搜寻到与其研究最为相关的录音整理稿或某个具体项目。把诸如主题指南、个人研究日记、结项报告(end of award reports)和出版物等重要文献集合在一起,建立用户指南。这类信息经田野调查和分析的主要研究者整理核对之后,无疑对将来的存档有很大的好处。

最后,定性数据的再使用和定性数据二次分析【Secondary Analysis of Qualitative Data】是一个自1990年代以来就激起争论的话题。和定量数据一样,定性数据也可以由新的分析者来处理——对旧数据提新问题、提供比较、证实或挑战已有的发现等。但是,我们应当始终意识到用于第二次分析的"旧数据"的不足之处。

首先,缺乏回答当前问题的准确数据会是一种令人沮丧的体验。其次,正如哈默斯利(Hammersley,1997)所强调的,在缺乏对研究过程的全面了解的情况下对数据质量和效度进行评价,比评估研究者自己的数据集要复杂得多。这是因为定性分析之路几乎从来不是线性的,在追究感兴趣的调查线索过程中,几乎总是会在一定程度上涉及试错法。

最后,尤其对定性数据而言,很难重构出访谈的场景或田野经验,而这通常在存档的访谈录音整理稿或田野笔记里都没有得到充分记录。其结果是,二手分析者无法了解最初的研究者所获取的那些不言而喻的知识,而这在某些情况下对数据的解读是至关重要的。在阅读录音整理稿之外,听听存档的访谈录音通常会帮助我们重构出访谈的场景。

——Louise Corti

(杨可译 沈崇麟校)

## 参考文献


Corti, L., Foster, J., & Thompson, P. (1995). *Archiving qualitative research data* (Social Research Update No. 10). Surrey, UK: Department of Sociology, University of Surrey.

Hammersley, M. (1997). Qualitative data archiving: Some reflections on its prospects and problems. *Sociology*, *31*(1), 131-142.

James, J. B., & Sørensen, A. (2000, December). Archiving longitudinal data for future research: Why qualitative data add to a study's usefulness. *Forum Qualitative Sozialforschung* [Forum: Qualitative Social Research] [online journal], *1*(3). Available: http://qualitative-research.net/fqs/fqs-eng.htm.

Mruck, K., Corti, L., Kluge, S., & Opitz, D. (Eds.). (2000, December). Text archive: Re-analysis. *Forum Qualitative Sozialforschung* [Forum: Qualitative Social Research] [online journal], *1*(3). Available: http://qualitativeresearch.net/fqs/fqs-eng.htm.


# ARIMA

ARIMA是自回归整合移动平均(Autoregressive, Integrated, Moving Average)的英文缩写,表示由博克斯和昭(Box & Tiao,1965,1975)首先提出,博克斯和詹金斯(Box & Jenkins,1976)进一步开发的广义随机【Stochastic】时间系列模型族的三个组成部分。估计单变量的ARIMA模型常被成功地用于预测和形成多变量分析的基础,如干预分析【Intervention Analysis】评估,或时间系列的Box-Jenkins建模法中的传递函数【Transfer Function】模型的主要基础。与传递函数模型一起使用的ARIMA模型,常与使用时间系列的回归【Regression】的模型作对比,但ARIMA模型也越来越多地与回归

模型连在一起使用。

## 三个组成部分

ARIMA 模型有三个组成部分，每一部分都描述了时间系列过程的一个主要性质。扼要地说，三个组成部分分别对应过去值（past values）效应，叫作自回归（AR）；过去的冲量（past shocks），叫作移动平均（MA）；存在的随机趋势，叫作整合（Ⅰ）过程。这三个部分将在下面参照一个时间系列过程的例子分别进行讨论。

### 自回归

时间系列过程经常受到它们自己的过去值的影响。我们来看一个有关公众赞同总统执政方式的百分比的问题。当前公众有关总统执政能力的意见，有一部分受到他们自己近来在这方面的态度的影响。更为一般地说，如果现在的值是过程的过去的值的函数，那么时间系列便是自回归的。过程过去值的效应用 ARIMA 模型的自回归部分，一般由下式表示：

$$y_t = a + b_1 y_{t-1} + b_2 y_{t-2} + \cdots + b_p y_{t-p} + \mu_t \quad (1)$$

每个观察值都由一个随机冲量和前面的观察值的线性组合组成。滞后值的数目决定了自回归的阶，并由 $p$ 给出，用AR($p$)表示。

### 移动平均

ARIMA 模型假设时间系列过程由一系列随机冲量驱动。随机冲量大致可被看成是一种随时间随机发生的投入的集合。更为专业地说，随机冲量是服从正态和恒等分布的随机变量，有零均值和常方差。考虑一下如下实例，即暴力犯罪的上报要求的变化。上报要求的变化会“冲击”暴力犯罪的统计值，因此报告的统计值就是此类冲量的函数。我们可以把这一函数用下式来表示：

$$y_t = a + \mu_t + c_1\mu_{t-1} + c_2\mu_{t-2} + \cdots + c_q \mu_{t-q} \quad (2)$$

式中的 $q$ 给出了过去冲量的数目，$\mu$ 影响着 $y$ 过程当前的值，在这里便是暴力犯罪的统计值。$\mu$ 的滞后值数目确定了移动平均的阶，它由 $q$ 给出，用 MA($q$)表示。

### 整合

ARIMA 模型假设过去的冲量和过去值的效应会随时间衰减：就冲量和赞同值而言，当今总统的赞同值，更多地反映了近来的，而不是时间较远的这两个值。例如，总统在他任期的早期犯的错误，随着时间的推移会被公众忘却。这就意味着，在广义 AR 和 AM 模型的组成部分中的参数【Parameter】值，必须受到一定的限制，以确保效应的衰减（即时间系列是平稳的）。不太严格地讲，平稳的时间系列的均值、方差和协方差，不会随时间的变化而变化。一个非稳定的时间系列叫作整合（Integrated）。某些经济学的时间系列（如经济增长）便被视为是整合的，因为它们展示的是增长的历史趋势。

在有可能进行分析之前，整合时间系列必须作一个转换。最为普遍使用的转换【Transformation】叫作差分化，凭借着这一变换，我们分析的是 $y$ 的变化，而非 $y$ 的水平（我们分析的是 $y_t - y_{t-1}$）。一个时间系列被差分至稳定的次数给出了整合的阶，以I($d$)表示。$d$ 可以是任何实数（如果 $d$ 是一个分数，那么这一过程便叫作分数化整合，而这样的模型则叫作 ARFIMA）。如果时间系列在水平上，而非在变化上整合，那么整合过程的阶 $d=1$。

AR(1)过程的稳定条件是 $b_1$ 必须在-1和+1 之间。如果 $b_1$ 不在这一范围内，这一过程的过去值的效应便会随时间累积，而连续的 $y$ 将趋向无限，也就是说，系列将不稳定，因而在（再）考虑作为过程的特征的自

回归的阶之前，可能还需要差分。在MA(1)这一情形中，需要将 $c_1$ 限定在-1和+1之间。在更高阶的过程中，为了确保稳定性，我们也必须对参数实施同样的限定。

## 广义模型

广义的 ARIMA $(p, d, q)$ 模型将上面给出的部分联合在一起，被写作如下形式：

$$y_t = a + b_1 y_{t-1} + b_2 y_{t-2} + \cdots + b_p y_{t-p} + \mu_t + c_1 \mu_{t-1} + c_2 \mu_{t-2} + \cdots + c_q \mu_{t-q} \quad (3)$$

为了产生平稳性，式中的 $y$ 已被差分过了适当次($d$，一般为0或1)，$p$ 给出了AR的滞后数，$q$ 则是MA的滞后数。任何时间系列可含AR、I和MA参数的任何数(包括0)，但参数的数目则将趋小。一般来讲，简单结构构成了时间系列过程，大多数过程都可作为为数不多的几个过去的冲量和过去值的一个函数而被清楚地解释。实际上，许多社会科学的过程，其性状就像一些简单的ARIMA(1,0,0)过程。

重要的问题是我们应该明白，一个给定的时间系列过程也许最好被视为一个ARIMA(0,0,0)过程，即所谓的白噪【White-Noise】过程。白噪过程是一个随机冲量系列。它没有自回归、整合或移动平均部分。在估计回归模型时，只有在误差是一个白噪过程时，推论才是正当的。

表示一个特定的系列的ARIMA模型，是通过一种特殊的方法学认定的。尤其是我们所关心那个过程的ARIMA模型，是一个在经验驱动的确认、估计和诊断的过程中，从一个观察的时间系列建立的(有关如何做到这一点的解释，请参见Box-Jenkins建模)。至于ARIMA模型和Box-Jenkins法的更为一般的处理方法，可参见麦克利里和海(McCleary & Hay，1980)的著作。

——Suzanna De Boef

(沈崇麟译校)

## 参考文献


Box, G.E.P., & Jenkins, G.M. (1976). *Time series analysis: Forecasting and control*. San Francisco: Holden-Day.

Box, G.E.P., & Tiao, G.C. (1965). A change in level of nonstationary time series. *Biometrika*, *52*, 181-192.

Box, G. E. P., & Tiao, G. C. (1975). Intervention analysis with applications to economic and environmental problems. *Journal of the American Statistical Association*, 20, 70-79.

McCleary, R., & Hay, R. A., Jr. (1980). *Applied time series analysis for the social sciences*. London: Sage.


# 阿罗不可能定理(Arrow's Impossibility Theorem)

肯尼思·阿罗(Kenneth J. Arrow，1921— )是20世纪最重要的经济学家之一。他对社会科学最重要的贡献就是阿罗不可能定理(Arrow，1963)。阿罗用一个以社会中的个人选择偏好为基础的社会考察了社会选择的逻辑，并认定了一个社会应该具备的用于社会选择规则的若干合理条件。不可能定理指出，当存在三个或三个以上选择方案时，不可能找到一种社会选择规则，总是能以一种满足这些理想条件的方式将个人偏好聚集成社会偏好。因此，即使每个个人的偏好完全一致且符合逻辑，但作为一个整体，社会偏好也仍然可能是不合逻辑的。

首先，假设社会中的每个个人在一系列选择方案中都有自己的偏好。一个社会选

择的规则是,考虑到所有个人的偏好,并将它们结合起来确定整个社会偏好的规则。例如,个人可能对于政治候选人有不同偏好,那么我们就要找出一个社会选择规则(例如少数服从多数原则)来结合人们的偏好,并决定谁应该当选。阿罗认定了社会选择规则应当满足的五个理想条件:

- 完备的。社会选择规则应当提供一个所有选择的完整排序。也就是说,对任意的选择 A 和 B 而言,社会选择规则应当告诉我们,究竟是 A 胜过 B,还是 B 胜过 A,或者在 A 和 B 之间是不存在社会差异的。
- 帕累托最优的。如果每个人在 A、B 之间都更偏好 A,那么社会选择规则应当将 A 排在 B 之上。在这种情况下,选择方案 A 而非方案 B 应当是一个帕累托最优的行动(一种使所有个体得益的行动),因此社会选择规则应当将 A 排在 B 之上。
- 可传递的。如果社会选择规则将 A 排在 B 之上,且将 B 排在 C 之上,那么 A 的排序应该高于 C。
- 无关方案互相独立的。A 相对于 B 的排序应当独立于其他的方案偏好。
- 非独裁的。社会选择规则应当不取决于单独某个人(独裁者)的偏好。

阿罗不可能定理的论证证明,不可能在有三个或以上选择方案的时候,找到一个社会选择规则总能满足这五个条件。也就是说,不管一个社会采用什么规则来确定其社会偏好,它要么具有某些不受欢迎的"非理性"特征,要么就完全是由一个人决定的。

不可能定理的结果对于那些研究民主选择的人来说是一个困扰。尽管如此,放宽不可能定理强加的某些条件,或者添加一些附加的限制条件,的确能使政治过程中的社会选择成为可能。一个重要的例子是,当假设各种方案按照某种一维谱排列,且个人对这些方案的选择偏好都是单峰的时候,就意味着偏好的强度在最受偏爱的方案中达到峰值,且与最受偏爱的方案距离越远,强度越小。在这个情况下,当人数为奇数时,少数服从多数原则会导致占中位数的投票者的偏好被采用(例如,Black, 1958; Downs, 1957)。社会规范和制度也可以提供一些办法来避免未知定理可能带来的影响(例如,Riker, 1980)。

——Garrett Glasgow

(杨可译 沈崇麟校)

## 参考文献


Arrow, K. J. (1963). *Social choice and individual values* (2nd ed.). New Haven, CT: Yale University Press.

Black, D. (1958). *The theory of committees and elections*. Cambridge, UK: Cambridge University Press.

Downs, A. (1957). *An economic theory of democracy*. New York: Harper Collins.

Riker, W. H. (1980). Implications from the disequilibrium of majority rule for the study of institutions. *American Political Science Review*, *74*, 432-446.


# 研究过程中的人为效应(Artifacts in Research Process)

研究中的人为效应这个术语指的是不受控制,并非故意的系统性偏差,它会影响研究结论的内在效度【Internal Validity】或外在效度【External Validity】。在社会科学研究中有存在"研究中的人为效应"的可能,因为人类参与者是有感情的、主动的生物体,而

不是被动、消极的客体。正因如此，在研究者和研究施行者（the agents of study）之间的互动也可能以一种研究者未曾察觉的微妙方式潜在地改变实验环境。

社会科学研究中的人为效应可能引起的潜在问题，在 19 世纪末 20 世纪初彰显了出来。举例来说，西电公司 1920 年代在伊利诺伊的霍桑工厂进行的一系列有争议性的研究，被解读成仅仅观察一个人的行为就足以改变其行动。在此之前，著名的德国心理学家奥斯卡·芬格斯特（Pfungst，1911/1965）20 世纪初在对一匹名为“聪明汉斯”的马的研究中，戏剧性地证明了人们的期望是如何在潜意识的层面上影响他人的行为的。在另一项经典的研究工作中，索尔·罗森茨魏希（Rosenzweig，1933）提出，实验环境本身应当被作为一个心理问题来加以考虑。他认为，我们必须考虑一种可能，即实验者和研究的参与者的态度和动机有可能对研究环境发生微妙的影响。尽管他们对这些研究中人为效应的潜在问题发出了警示信号，但直到 1950 年代末，1960 年代初，这个问题及其对社会科学研究的深远影响才得到了系统的检视。这一研究成果首先由罗森塔尔和罗斯诺（Rosenthal & Rosnow，1969）结集出版，通常关注的是人为效应的两个潜在来源：研究者和研究的参与者。

## 与研究者有关的人为效应

因为研究者不是科学研究过程中被动、冷漠的参与者，他们有时候会不知不觉地在无意中作出一些会对研究设计产生人为影响的行为。罗森塔尔（Rosenthal，1966）在一部重要著作中定义了两种广义的，与研究者有关的人为效应类型：无互动的人为效应和互动的人为效应。无互动的人为效应是不会对研究参与者的行动产生实际影响的偏误。更确切地说，是研究者观察、阐释或者报告研究结果的方式所产生的系统性的偏差。例如，研究者的观察可能偏向于支持研究者的假设。同样，研究者阐释数据的方式也可能偏向于支持其关注的研究假设。尽管这两种无互动的偏误可能是无意识的，但第三种则肯定是有意而为之。研究者可能会编造或操控数据，以便为其假设提供支持，这当然是不符合职业道德的。

第二种与研究者相关的人为效应，则与研究者和研究参与者之间的互动有关。例如，研究者的生物社会特征（例如民族、性别、年龄）有在无意识的层面对参与者的行为发生影响的潜在可能。同样，研究者的心理特征（例如人格）也可能造成参与者以一种看似与被调查项目无关的方式作出反应。情境效果（例如研究者在研究过程中的行为变化）也会改变人们对实验刺激的反应方式。最后，正如罗森塔尔（Rosenthal，1966）所证明的，仅仅是对于研究假设的了解也会影响研究者的行为，导致自证预言或实验者期望【Experimenter Expectancy】偏误，这是目前研究最为深入的一种互动的人为效应的来源。

## 参与者的人为效应

研究参与者也是积极的、有目的的当事人，他们了解自己作为研究对象的角色。马丁·奥恩（Orne，1962）指出，尤其是那些被帮助科学事业的愿望促动的自愿受试者【Volunteer Subjects】会扮演“好被试者”的角色。在这个位置上，他们努力寻找提示线索（需求特征【Demand Characteristics】）并对其作出反应，表面上这样做会令其行为更可能支持研究者的假设。但是米尔顿·罗森堡（Rosenberg，1969）还指出，评价焦虑【Evaluation Apprehension】也会促动研究参与者的行为。他提出，由于他们担心被负面评价，导致他们倾向于对实验

刺激作出大概能让他们“看上去不错”的反应。研究环境中的参与者行为的第三种可能动机是服从研究者的愿望,因为后者被视为权威人物。不管背后的动机是什么,其后果是人们对研究中的实验刺激作出了系统性的有偏反应。

## 减少研究中的人为效应

对研究中的人为效应潜在来源的确认工作向我们提示了探查人为效应存在的方法或减少将其带入研究设计【Research Design】的方法。例如,潜在的无互动的研究者有关的人为效应,可以通过独立的重复【Replication】实验结果,以及向其他科学家开放研究数据加以认定。罗森塔尔(Rosenthal,1966)记述了如何采用期望控制设计来探测实验者期望偏误是否存在。标准化的研究规程、实验设计中的双盲程序【Double-Blind Procedure】以及实验处理和处理效果测量的时空分离则代表了减少研究者相关的人为效应的其他一般性策略。

避免特定的参与者相关的人为效应的策略,包括引入替代性的任务取向线索、创设可以减轻评价焦虑或减少对需求特征作出反应的实验环境。使用不知道自己身在研究中的参与者会使得反应性【Reactivity】问题最小化,但这一策略会引起侵犯隐私【Privacy】的伦理问题。

奥恩(Orne,1969)描述了如何利用半控制的被试者来获取实际研究中的需求特征的办法。这些参与者基本上是作为研究者的“顾问”参与研究的。半控制的被试者在了解实验设计的各个方面之后,被要求谈论其行为背后的理念,或思考当其真的置身于实验刺激之下时会如何反应。最后,研究者确认和控制潜在的研究人为效应的最重要的机制,就是由不同的研究者对研究结果进行独立的重复研究。

——David B. Strohmetz
Ralph L. Rosnow
(杨可译　沈崇麟校)

## 参考文献

Orne, M. T. (1962). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist*, *17*,776-783.

Orne, M. T. (1969). Demand characteristics and the concept of quasi-controls. In R. Rosenthal & R. L. Rosnow (Eds.), *Artifact in behavioral research* (pp. 143-179). New York: Academic Press.

Pfungst, O. (1965). *Clever Hans: The horse of Mr. von Osten.* New York: Holt, Rinehart, & Winston. (Original work published 1911.)

Rosenberg, M. J. (1969). The conditions and consequences of evaluation apprehension. In R. Rosenthal & R. L. Rosnow (Eds.), *Artifact in behavioral research* (pp. 279-349). New York: Academic Press.

Rosenthal, R. (1966). *Experimenter effects in behavioral research.* New York: Appleton-Century-Crofts.

Rosenthal, R., & Rosnow, R. L. (Eds.). (1969). *Artifact in behavioral research.* New York: Academic Press.

Rosenzweig, S. (1933). The experimental situation as a psychological problem. *Psychological Review*, *40*, 337-354.

Rosnow, R. L., & Rosenthal, R. (1997). *People studying people: Artifacts and ethics in behavioral research.* New York: W. H. Freeman.

Strohmetz, D. B., & Rosnow, R. L. (1994). A mediational model of research artifacts. In J. Brzezinski (Ed.), *Probability in theory-building: Experimental and nonexperimental approaches to scientific research in psychology* (pp. 177-196). Amsterdam: Editions Rodopi.

## 人工智能(Artificial Intelligence)

人工智能(AI)的主要目的是研究如何建立智能系统。AI尤其关注如何用智力行为,如问题的求解、推理和学习来开发计算机程序。人工智能也因此而得名。

自1965年人工智能一词首次出现以来,AI一直呈起伏不定的周期状。最初,解决电子游戏中遇到的各种问题一直是AI研究的重要课题。1997年,IBM的深蓝国际象棋下棋系统(IBM's Deep Blue Chess System)打败了世界国际象棋冠军加里·卡斯帕罗夫。专家系统也已经被成功地开发出来,其在操作应用上已经达到在真实世界里专家的水平。将在特定领域的人类专家的知识库中汲取的知识用于机器推理,并将它广泛用于科学、医学、商业和教育等各种领域。机器学习的崛起则成为AI研究中的又一个重要课题,人们致力于建立有学习能力的人工智能系统。仿照生物进化的图形,基因算法催生了数量众多的彼此竞争的候选解决方案,而众多的竞争方案有助于更好的解决方案的诞生。基于代理人的模型(Agent-based model)证明,普遍的智能行为(globally intelligent behavior)可以出现在较小的结构的互动中。基于个体代理人的互动、智能可被视为社会特性的体现,而不仅仅是个体代理的一种性质。近年来,AI一直在转向构建真实环境中的智能代理人,如网上拍卖。得益于对一个在有着不同的能力、目标和信仰的多个代理环境中的人类社会的抽象,多代理人系统成为聚合各种AI技术的平台。

人们也一直在把AI路数用于社会、心理认知和语言现象。在理论构建层次,AI技术一直在被用于理论的定型,以及社会和组织的模拟。例如,一个博弈论【Game Theory】问题——"囚徒"困境,这是一个经典问题。对这一问题的研究便得益于计算机模拟【Simulation】。它使我们能对在什么样的条件下,那些有动机违背协议的利己的行动者之间,有可能至少在短时间内有可能进行合作这一问题进行探讨。近来,多代理人系统已成为一种给社会生活建模的新方法。这些模型显示,信息是以什么样的模式传播的,规范是如何出现的,约定是如何协调的,而集体行动的参与问题则可被看成是那些对受到的某一影响作出反应时,那些相互影响的适应性代理人(adaptive agents)之间互动出现的行动。

在数据分析层次,AI技术一直被用于使用符号处理器智能地分析定性数据(用那些诸如民族志这样的方法收集的数据)。而在定量数据分析中,AI技术则大大提升了统计程序的能力。例如,已经证明,我们可以立即用神经网络【Neural Networks】替代,甚至胜过多元回归及其他多变量方法。最近,知识发现的过程显示,社会科学研究有着巨大的潜力。通过AI技术的使用,我们可以对数据作进一步发掘(从数据中发现趋势或范式的过程),进而将它们转化为有用和能解释的知识。

——Xiaoling Shu

(沈崇麟译校)

### 参考文献

Alonso, E.(2002). AI and agents: State of the art. *AI Magazine*, *23*(3), 25-29.

Bainbridge, E. E., Carley, K. M., Heise, D. R., Macy, M. W., Markovsky, B., & Skvoretz, J.(1994). Artificial social intelligence. *Annual Review of Sociology*, *20*, 407-436.

Conte, R., & Dellarocas, C.(2001). Social order in info societies: An old challenge for innovation. In R.

Conte & C.Dellarocas(Eds.), *Social order in multi-agent systems*(pp.1-16). Boston: Kluwer Academic.

Luger, G.F.(2002). *Artificial intelligence: Structures and strategies for complex problem solving*(4th ed.). Reading, MA: Addison-Wesley.

Macy, M.W., & Willer, R.(2002). From factors to actors: Computational sociology and agent-based modeling. *Annual Review of Sociology*, *28*, 143-166.

## 人工神经网络(Artificial Neural Network)

参见神经网络【Neural Network】。

## 关联(Association)

变量之间关系(关联)的研究是社会研究的基本任务。两个(定类)变量的研究可以用双向的列联表【Contingency Table】通过行百分比、列百分比或总百分比的计算制成交互分类表来研究。如果变量是连续的,我们同样也可以采用这种方法。但必须将变量通过分组分出类别。两个变量之间的关联概括性量度(summary measure)由各种量度两个变量 $X$ 和 $Y$ 之间的关系强度,以及定类、定距和定比的变量的方向的关联系数(association coefficients)给出。方向可以是正的(一致的,高 $X$ 值有相应的高 $Y$ 值),也可以是负的(不一致,高 $X$ 值对应于低 $Y$ 值)。可区分两大类关联系数:

• 不对称量度【Asymmetric Measures】或关联系数,假定一或两个变量(如 $X$)可以被确定为自变量,而另外一个变量(如 $Y$)可以被确定为因变量。例如,变量 $X$ 可能是性别,而变量 $Y$ 可能是职业地位。

• 对称量度【Symmetric Measures】或关联系数不需要区别自变量和因变量。因而,只有在方向的区别是不可能的时候,它们才可以使用。例如,研究人员研究不同的闲暇活动(运动和看剧)之间的关系便是一例。然而,在自变量和因变量可以确定的时候,我们同样也可以计算对称系数。

下面是一些对称系数的例子(如 Healey, 1995; SPSS, 2002):phi($\phi$); Cramér $V$;定类变量的列联系数($C$); Kendall tau$_a$($\tau_a$),tau$_b$($\tau_b$)和 tau$_c$($\tau_c$); 定序变量的 Goodman 和 Kruskal gamma ($\gamma$); 等级数据(序)的 Spearman rho ($\rho$)和定距或定比变量的 Pearson $r$。表格的大小和两个变量的类别数决定了使用某一水平上的何种量度。定向的对称系数常常也被称作相关【Correlation】系数,有时,相关(correlation)一词专指两个连续变量之间的关联的。不对称系数的例子(如 Healey, 1995; SPSS, 2002)有 lambda ($\lambda$), Goodman 和 Kruskal $\tau$;定类变量的不确定系数(uncertainty coefficient);定序变量的 Somers $d$ 和因变量为定距或定比变量,自变量为定类变量的 eta ($\eta$)。更进一步的对称系数,如对称 lambda 和对称 Somers 可以通过均值的平均从不对称系数导出。

上面提到的系数(除 $\eta$ 外)都假定两个变量有相同的量度水平。如果分析的变量有不同的量度水平,那么我们建议先降低量度水平,然后在较低量度水平上计算恰当的关联量度。遗憾的是,我们常常需要这样做,因为在标准的统计软件里,一般都没有适用于混合量度水平的系数。混合量度水

平的系数可从丹尼尔的广义相关系数中推导出来，例如（Daniels，1944；Kendall，1962，pp. 19-33）。

关联的概念也可广泛地应用于多元分析。多变量关联系数的例子包括两组变量 $X=\{X_1, X_2, \cdots, X_p\}$ 和 $Y=\{Y_1, Y_2, \cdots, Y_p\}$ 之间的典型相关分析【Canonical Correlation Analysis】；一组变量 $X=\{X_1, X_2, \cdots, X_p\}$ 变量和 $Y=\{Y_1\}$ 的多重相关【Multiple Correlation】系数 $R$（来自多元回归）；和来自双变量分析【Bivariate Analysis】的两个定类变量 $X$ 和 $Y$ 之间的惯性（inertia）等。

一种用于双变量和多变量列联表的备择策略，是基于对数线性模型【Log-Linear Model】的关联模型【Association Models】。对数线性模型使我们得以设定和分析不同的、复杂的关联模式，而关联系数则给出了概括性量度。

关联系数测量两个变量之间的关系强度。人们常常要求我们提供解释强度的指南，但是提供一般性的指南是不可能的。关联强度取决于样本的同质性程度、变量的可信度和两个变量之间的关系类型。在心理实验中（高同质性样本、高信度变量、直接的因果关系），0.4 的系数可能表示显著的弱关系，而这同样的值，在社会学调查中（异质性样本、低可信度、非直接因果关系）则可有保留地被视为强关系。因此，在确定阈值时，重要的是在考虑这些因素的同时，考虑前面的研究结果。

——Johann Bacher

（沈崇麟译校）

## 参考文献

Daniels, H.E. (1944). The relation between measures of correlation in the universe of sample permutations. *Biometrika*, *35*, 129-135.

Healey, J.F. (1995). *Statistics: A tool for social research* (3rd ed.). Belmont, CA: Wadsworth.

Kendall, M. G. (1962). *Rank correlation methods* (3rd ed.). London: Griffin.

SPSS. (2002, August). Crosstabs. In *SPSS 11.0 statistical algorithms* [online]. Available: http://www.spss.com/tech/stat/algorithms/11.0/crosstabs.pdf.

# 关联模型（Association Model）

虽然关联【Association】一词已经广泛使用，但关联模型一词在有关分类数据分析【Categorical Data Analysis】的文献中却有着特定的含义，指一种拟合交互分类表中的观察频数的统计模型族。拟合的目的是测量两个或更多个有序分类变量之间的关联强度。双向表测量的是组成交互分类表的两个定类变量之间的关联强度。在三向或更多向的表中，任何组成交叉分类表的有序的分类变量对之间关联强度，都是可以量度的。尽管有些关联模型使用了分类的先验的排序，但其他模型却并非始于诸如这样的假定，而是通过估计来确定分类的排序。关联模型是对数线性模型【Log-Linear Model】或对数双线性模型的特例。

里奥·戈德曼对关联模型的开发居功至伟。他于 1979 年发表在《美国统计协会学刊》（*Journal of the American Statistical Association*）的论文是这一领域发展的奠基石。这篇如种子般的论文及其他有关文章，都收录在他于 1984 年出版的《有定序类别交叉分类数据的分析》（*The Analysis of Cross-Classified Data Having Ordered Categories*）这本书中。在讨论扩展的三向或更多向的表之

前,我先来介绍双向表的典型情形。我将用三个社会学和人口学的例子,来形象地介绍关联模型的用处。

## 双向交叉分类表的一般设置

对 $R$ 和 $C$ 的双向表中第 $i$ 行和 $j$ 列的格 ($i=1,\cdots,I;j=1,\cdots,J$),令 $f_{ij}$ 和 $F_{ij}$ 分别为某一模型中的观察频数和期望频数。我们可在不失其普遍性的前提下,将一个对数线性模型写为如下的形式:

$$\log(F_{ij})=\mu+\mu_i^R+\mu_j^C+\mu_{ij}^{RC} \quad (1)$$

该式中的 $\mu$ 是主效应, $\mu^R$ 和 $\mu^C$ 分别是行效应和列效应,而 $\mu^{RC}$ 则表示期望频数的对数的交互效应。方程(1)中的所有参数都受 ANOVA 型的正态化约束(参见 Powers & Xie, 2000, pp. 108-110)。我们常常令 $\mu^R$ 和 $\mu^C$ 不受约束,并在正态化上对其进行估计。这一实用方法也叫作行和列变量的边际分布的"饱和(saturation)"。其所感兴趣的是 $\mu^{RC}$。在一种极端的情形中,因 $\mu^{RC}$ 可能全都为零而导致一个独立模型。在另一种情形中, $\mu^{RC}$ "饱和"有 $(I-1)(J-1)$ 个自由度,得到的是精确预测(exact prediction)(对所有的 $i$ 和 $j$, $F_{ij}=f_{ij}$)。

在一般情况下,研究者的兴趣在于通过调整 $\mu^{RC}$ 的设定,在两种极端情形之间来拟合模型。我们很容易看到双向表中的所有的优势比【Odds Ratio】都是交互参数($\mu^{RC}$)的函数。令 $\theta_{ij}$ 为一个 2 × 2 的子表的局部概率比率的对数,该子表来自母表的两个比邻行类和两个比邻列类的四个比邻的格:

$$\theta_{ij}=\log\{[F_{(i+1)(j+1)}F_{ij}]/[F_{(i+1)j}F_{i(j+1)}]\},$$
$$i=1,\cdots,I-1;j=1,\cdots,J-1$$

我们假定行和列变量是在某些尺度上的定序变量 $x$ 和 $y$。这些尺度可以是观察的也可以是潜在的。那么一个线性关联模型就可以写作:

$$\log(F_{ij})=\mu+\mu_i^R+\mu_j^C+\beta x_i y_j \quad (2)$$

式中的 $\beta$ 是测量两个分别表示行和列变量的尺度 $x$ 和 $y$ 之间的关联的参数。如果两个尺度 $x$ 和 $y$ 是直接观察到的,或从内部的资源估算的,那么方程(2)便可直接使用对数线性模型的最大似然估计【Maximum Likelihood Estimation】。

## 双向表关联模型

如果我们没有两个尺度 $x$ 和 $y$ 的额外信息,那么我们既可以自行对尺度作出假定,也可以根据情况对它们进行估计。不同的做法将会产生不同的关联模型。下面我们一起来评议一下其中最重要的几个。

## 一致关联(格式)

如果变量的类有正确的排序,研究者可以作一个简约的假定,即假定排序的位置形成了一个尺度(如 $x_i=i$, $y_j=j$)。我们把这种做法叫作整数计分(integer scoring)。整数计分简化后得到如下的一致关联模型:

$$\log(F_{ij})=\mu+\mu_i^R+\mu_j^C+\beta_{ij} \quad (3)$$

研究者可以用实际的数据来估计这一模型,看看这个假定是否站得住脚。

## 行效应和列效应模型

虽然一致关联模型是基于行和列二者都是整数计分这一情形的,但是研究者可能希望只是行或列是整数计分的。在整数计分只用于列变量时,得到的模型叫作行效应模型(row effect model)。以此类推,在整数计分只用于行变量时,得到的模型叫作列效应模型(column effect model)。以行模型为例,我们可从方程(2)导出如下的模型:

$$\log(F_{ij})=\mu+\mu_i^R+\mu_j^C+j\phi_i \quad (4)$$

这一模型之所以叫作行效应模型，是因为潜在的行变量的计分($\phi_i = \beta x_i$)是在我们将整数计分应用于列变量之后估计得到的。这就是说，$\phi_i$是在行和列变量之间的关联上的“行效应”。注意，这里的行效应(row effect)和列效应(column effect)这两个词，其含义不同于$\mu_i^R$和$\mu_j^C$。后者拟合的是行和列变量的边际分布的饱和度。

## Goodman 关联模型

研究者可以再进一步将行和列的计分都作为未知的来处理。有两个 Goodman 关联模型被设计用来估计诸如这样的潜在计分(Goodman，1979)。Goodman 关联模型 I 将方程(1)简化为如下的形式：

$$\log(F_{ij}) = \mu + \mu_i^R + \mu_j^C + j\phi_i + i\varphi_j \quad (5)$$

式中的$\phi_i$和$\varphi_j$与行和列效应模型一样，分别表示未知的行和列计分。然而，我们必须增加三个正态化的约束，以唯一地确定$\phi_i$和$\varphi_j$的$(I+J)$个未知参数。

因为使用了整数计分，所以 Goodman 关联模型 I 要求，如方程(5)所示的那种行和列变量的先验正确排序。这一要求意味着，模型对行和列变量的类别位置变化不具有不变性。如果研究者不知道类别的正确排序，或实际上需要确定类别的正确排序，使用模型 I 便是不恰当的。正因如此，Goodman 关联模型 II 才更为人们所关注。这一模型的形式如下：

$$\log(F_{ij}) = \mu + \mu_i^R + \mu_j^C + \beta\phi_i\varphi_j \quad (6)$$

式中的$\beta$是关联参数，而$\phi_i$和$\varphi_j$是需要估计的未知计分。$\phi_i$和$\varphi_j$同样受四个正态化约束，因为它们每一个都要求位置和尺度二者的正态化。

正如方程(6)所示，Goodman 关联模型 II 交互部分($\mu^{RC}$)是未知参数的乘积——对数双线性设定。这个模型叫作对数双线性模型(log-multiplicative model)或者可简单地叫作 $RC$ 模型。$RC$ 模型是很有用处的，因为它使我们得以在即使不知道行和列变量的正确排序时，也可以估计未知参数。我们需要做的全部事情，就是假设定序尺度的存在。模型可以通过估计来揭示顺序。

表 1 是上述关联模型的概括比较。第二列显示模型的交互参数($\mu^{RC}$)的设定。每个$\mu^{RC}$设定的自由度数则在第三列列出($DF_m$)。如果没有其他的参数需要估计，那么模型的自由度$DF_m$便等于$(I-1)(J-1)$。计算局部概率比率(local log-odds ratio)的计算公式列在了最后一列。

**表 1　关联模型比较**

| 模型 | $\mu^{RC}$ | $DF_m$ | $\theta_{ij}$ |
|---|---|---|---|
| 一致关联模型 | $\beta_{ij}$ | 1 | $\beta$ |
| 行效应 | $j\phi_i$ | $(I-1)$ | $\phi_{i+1}-\phi_i$ |
| 列效应 | $i\phi_j$ | $(J-1)$ | $\phi_{j+1}-\phi_j$ |
| 关联模型 I | $j\phi_i+i\phi_j$ | $I+J-3$ | $(\phi_{i+1}-\phi_i)+(\phi_{j+1}-\phi_j)$ |
| 关联模型 II ($RC$) | $\beta\phi_i\phi_j$ | $I+J-3$ | $(\phi_{i+1}-\phi_i)(\phi_{j+1}-\phi_j)$ |

Goodman 关联模型 II ($RC$ 模型)很容易被扩展到多个潜在维度，这样$\mu^{RC}$就变成了：

$$\mu_{ij}^{RC} = \sum \beta_m \phi_{im}\varphi_{jm} \quad (7)$$

式中的求和符号涉及所有可能的 $m$ 维，而参数则必须受正态化的约束。这样的模型叫作 $RC(M)$ 模型。详细内容请参见戈德曼的有关著作(Goodman，1986)。

## 三向和更多向表的关联模型

下面我们将主要来讨论三向表问题。

我们很容易将问题推广到三向表。令 $R$ 和 $C$ 分别表示行和列，$L$ 表示层，并用 $i$ ($i=1,\cdots,I$)，$j$ ($j=1,\cdots,J$) 和 $k$ ($k=1,\cdots,K$) 分别表示它们的类别。在一般的研究设置中，研究者的兴趣在于，在层 $L$ 上的 $R$ 和 $C$ 之间的关联。例如，在趋势分析中，$L$ 可能代表不同的年或同期群（cohorts）。在比较研究中，$L$ 可能代表不同的国家或群体。这就是说，研究者关注的是不同层次上的 $R$ 和 $C$ 的关联模式。

令 $F_{ijk}$ 为在第 $i$ 行、$j$ 列和 $k$ 层的期望频数。饱和对数线性模型（saturated log-linear model）可写为如下的形式：

$$\log(F_{ijk}) = \mu + \mu_i^R + \mu_j^C + \mu_k^L + \mu_{ij}^{RC} + \mu_{ik}^{RL} + \mu_{jk}^{CL} + \mu_{ijk}^{RCL} \quad (8)$$

在一个典型的研究设置中，我们的主要兴趣集中在不同层次的 $RC$ 关联的变差。因而基线（对虚无假设）便是下面这样的条件独立模型（conditional independence model）：

$$\log(F_{ijk}) = \mu + \mu_i^R + \mu_j^C + \mu_k^L + \mu_{ik}^{RL} + \mu_{jk}^{CL} \quad (9)$$

这就是说，为了了解特定层的 $RC$ 关联，研究者必须设定和估计 $\mu^{RC}$ 及 $\mu^{RCL}$。

有两种用得比较广泛，将关联模型推广到三向或更多向表的方法。第一种是先为典型的 $R$ 和 $C$ 之间关联模式设定一个关联模型，然后估计不同层的参数，或检验不同层的参数是否不同（Clogg, 1982a）。一般，这一方法是依据 $RC$ 模型来设定 $\mu^{RC}$ 和 $\mu^{RCL}$，这样我们就需要把方程（8）修改成如下的形式：

$$\log(F_{ijk}) = \mu + \mu_i^R + \mu_j^C + \mu_k^L + \mu_{ik}^{RL} + \mu_{jk}^{CL} + \beta_k \phi_{ik} \varphi_{jk} \quad (10)$$

这就是说，受制于模型的设定和统计检验，参数 $\beta$、$\phi$ 和 $\varphi$ 是针对每一层的。研究者也有可能对各个层进行检验（如一致关联、列效应和行效应模型），式中的参数 $\phi$ 和/或 $\varphi$ 是作为整数计分插入的，而非估计的。

第二种方法叫作对数乘积多层效应模型（log-multiplicative layer-effect model），或称为“Unidiff 模型”，它允许对 $R$ 和 $C$ 之间的典型关联作灵活的设定，然后将它的跨层变差约束为对数乘法的（log-multiplicative）（Xie, 1992）。这就是说，我们给 $\mu^{RC}$ 以灵活的设定，但却对 $\mu^{RCL}$ 加以约束，这样方程（8）便变成了如下形式：

$$\log(F_{ijk}) = \mu + \mu_i^R + \mu_j^C + \mu_k^L + \mu_{ik}^{RL} + \mu_{jk}^{CL} + \phi_k \psi_{ij} \quad (11)$$

使用第二种方法，$RC$ 关联便不再限于一个特定的模型，且的确可以用 $(I-1)(J-1)$ 个虚拟变量饱和。而其特例，$R$ 和 $C$ 之间的典型关联模型就是 $RC$ 模型，这两个模型共存，则会导致三向 $RCL$ 对数乘法模型。鲍尔斯和谢宇（音译）（Powers & Xie, 2000, pp. 140-145）给我们提供了有关第二种方法的变差和实际含义更为详尽的讨论。注意，这两种方法都是由戈德曼（Goodman, 1986）提出的，由戈德曼和胡特（Goodman & Hout, 1998）进一步扩展的一般框架的特例。

## 应用

关联模型在社会研究中得到了广泛的应用。下面将给出一些具体的例子。第一个例子关乎量度问题。关于这一例子的详尽讨论请参见克洛格的著作（Clogg, 1982b）。克洛格的目的在于为一个测量人们对流产的态度的定序变量制订测量尺度。这一变量是根据 Guttman 尺度构建的，但那些不符合这一尺度的回答模式的个案，则都被归入了单独的一类，即“错误的回答”类。在一般情况下，量度都需要一种“工具”，在这一例子中，克洛格使用

了一种来自同一调查的婚前性行为的态度量度。其潜在的假设是对流产的态度的尺度可能从它对婚前性行为的态度的关联得见一斑。克洛格用对数乘法模型来估计两个变量的不同类别之间的关联的计分。注意,对数乘法 *RC* 模型假定类别是定序的,但未必是正确排序的。这样,估计在揭示尺度的同时,也揭示了排序。通过估计,克洛格揭示了比邻类别间的距离是不等的,而给出"错误回答"的都是对流产持中间态的回答人。

第二个例子是有关对数乘法分层效应模型在跨国代际流动问题研究中的应用(Xie, 1992)。它的基本思想是在允许和检验父子的职业(如 $\psi_{ij}$)之间的双向关联的不同的参数化的同时,强行对分特定层参数的跨国的差异进行概括[如方程(11)的 $\phi_k$]。这样参数 $\phi_k$ 便被用来表示不同社会的社会开放性或封闭性。

第三个例子涉及人类生育问题研究,它的基本设置不是对数线性,而是对数比率,从这个意义上讲,它是非常规的。数据结构是由一张用年龄和国家做交叉分类的频数表(出生)和一张对应的关联暴露(妇女-年)表组成的。两个表之间的比率产生了特定国家和特定年龄的生育率。统计建模的目的是依据生育水平和生育控制粗略地描述生育的年龄模式。在常规的人口学研究中,这个问题通常都用科尔和特拉塞尔的 *Mm* 法来处理的。谢和皮门特尔(Xie & Pimentel,1992)证明,这一方法与对数乘法分层效应模型是等价的。用出生作自变量,而暴露(exposure)作为一种"补偿(offset)"。这样,科尔和特拉塞尔法中的参数 *M* 和 *m* 便可和模型中的其他未知参数一起在统计上估计。

## 估计

用一致、行效应、列效应和 Goodman 关联模型Ⅰ,这些模型的估计是一目了然的。用户可以用任何一个估计对数线性模型的计算机软件来做。而比较复杂的则是在 *RC* 交互采取未知参数的积这种形式时,即做对数乘法或对数双线性设定时。这时,需要进行迭代估计。它的基本原理是在估计一组参数时,轮流将另一组未知参数看作已知,反复进行直至二者都趋于稳定。一些专门的计算机软件,如 ASSOC 和 CDAS,GLIM 和 STATA 都有 PC 版可供我们使用。对所有真心想使用关联模型的读者,我则推荐一种叫 Lem 的软件,这个软件在使用灵活的同时,还可以估计对数乘法模型的不同形式。关于计算机子程序和特殊程序更新的信息,请参见我的个人网站。

——Yu Xie

(沈崇麟译校)

## 参考文献

Clogg, C. C. (1982a). Some models for the analysis of association in multiway cross-classifications having ordered categories. *Journal of the American Statistical Association*, *77*, 803-815.

Clogg, C. C. (1982b). Using association models in sociological research: Some examples. *American Journal of Sociology*, *88*, 114-134.

Goodman, L. A. (1979). Simple models for the analysis of association in cross-classifications having ordered categories. *Journal of the American Statistical Association*, *74*, 537-552.

Goodman, L. A. (1984). *The analysis of cross-classified data having ordered categories*. Cambridge, MA: Harvard University Press.

Goodman, L. A. (1986). Some useful extensions of the usual correspondence analysis approach and the usual log-linear models approach in the analysis of contingency tables. *International Statistical Review*, *54*, 243-309

Goodman, L. A., & Hout, M. (1998). Understanding the Goodman-Hout approach to the analysis of

differences in association and some related comments. In Adrian E. Raftery (Ed.), *Sociological methodology* (pp. 249-261). Washington, DC: American Sociological Association.

Powers, D. A., & Xie, Y. (2000). *Statistical methods for categorical date analysis*. New York: Academic Press.

Xie, Y. (1992). The log-multiplicative layer effect model for comparing mobility tables. *American Sociological Review*, 57, 380-395.

Xie, Y., & Pimentel, E. E. (1992). Age patterns of marital fertility: Revising the Coale-Trussell method. *Journal of the American Statistical Association*, 87, 977-984.

# 假定(Assumptions)

在社会科学中,假定是无所不在的。理论工作都从公理或假定开始。这些假定产生了涵盖广阔领域的可检验的研究命题。在经验工作中,统计程序一般都内含各种各样的假定,涉及变量的量度性质、分布形式和不可观察的操作(如同方差性【Homoscedasticity】或误差的正态分布【Normal Distribution】)。经验工作的假定的讨论都在那些特定方法(如普通最小二乘【Ordinary Least Squares】)的词条中进行,而在这里我们则主要在理论层面上来讨论假定。

特定理论【Theory】的目的在于产生涉及可观察的现象之间可检验的寓意(implications)。理论的核心是一组假定。假定体现了波普(Popper, 1963)称之为关于自然的"猜想"——被检验的猜想,而根据牛顿的观点则是检验其逻辑的寓意。一组假定的基本性质是内在的逻辑一致性。此外,还有三个期望的理论性质相随:(a)它的一组假定应尽可能简短;(b)它的可观察的寓意应尽可能多种多样;(c)它的可观察的寓意应包括还未曾观察的现象或关系,即所谓新奇的预测。

因此,理论有两部分结构:一小部分含假定,而一大部分且始终在增长的部分包含寓意。图1给我们提供了一张直观的理论结构图。

一个理论可能满足上述三种性质,但它却是错误的。这就是说,人们可以发明一种想象的世界,建立一组它运作的简约假定,推论出各种各样的经验的结果,然而,通过经验检验却了解到,没有已知的世界,其运行与假定的世界的性质中推导出来的寓意相符。这就说明了经验分析为什么是不可或缺的,换句话说,只是理论本身是不能满足可靠的知识积累的。

关于这一术语我们作一个说明:假定是一个通用的术语,诚如我们所知,它在理论和经验著作中都使用。在理论著作中使用的术语,有时更为严格,包括了公理(axiom)等,它含有自明(self-evident)的意思,但假设(postulate)却没有这个意思。因而更适合猜想类的工作,推动人们有所发现。其他的术语,包括可供使用的有前提(premise),多姿多彩的启始原理(starting principle)和危险的假设(hypothesis),它们不仅被用作公设(像牛顿发明的假说-演绎法【Hypothetico-Deductive Method】那样),也被用作要被检验的可观察的命题。




**图 1　理论的结构**

假定从何而来？通常，分析专题领域的那些框架本身就包含着各种各样的关系和函数【Functions】，其中有些可提供丰富的假定。一般来讲，用数学式表达的函数可生成有意义的假定，因为：（a）它们特有的陈述有着相当的广度和深度；（b）它们可以用数学工具进行操作，从而有可能导出大量寓意。数学函数同语言文字表述的普通规律不同，因为它们能够导出新奇的预测。

导出推论的工具包括全套的数学工具。除了函数之外，它们还包括分布【Distribution】、矩阵【Matrices】、不等式和偏微分方程。例如，借助函数表达的数量关系，可用它们的分布来描述其特征，而分布的形式则形成了预测的特征。得到的结果可能令人惊讶不已。假定看起来似乎平淡无奇，但却可以以出人意料的方式涉足广泛的领域。举例而言，在将司法评估函数作为一种假定时，可预测父母的馈赠、母丧和父丧的差别、战斗老兵外伤后压力风险的差别、修道院和托钵修道会对社会贡献的差别、不平等对移民的影响（Jasso，2001）。

为了确认假定的有用性，简单易行的做法是，使一个特定理论的电子表格显示它的一组假定中那些被用于导出每一种预测的假定。这样的设计不仅可一目了然地看出是哪些假定的作用最大，还可使我们得知每种预测究竟需要多少假定。图 2 给我们提供的一个这种形式的直观设计，叫作默顿理论导源图。

理论：______________________　　　　日期：______________

页：______________

| 预测 | 假设 | | | | |
|---|---|---|---|---|---|
| | | | | | |
| | | | | | |
| | | | | | |
| | | | | | |
| | | | | | |
| | | | | | |
| | | | | | |
| | | | | | |
| | | | | | |
| | | | | | |
| | | | | | |
| | | | | | |
| | | | | | |
| | | | | | |
| | | | | | |
| | | | | | |
| | | | | | |
| | | | | | |
| | | | | | |
| | | | | | |
| | | | | | |

注：

**图 2　默顿理论导源图**

一个理论和它的假定一般都是比较基本的（基础研究【Basic Research】）。圣杯的发现是支配社会行为世界的终极力量的发现。而我们手头研究的理论和假定，即使最有成果的和进行了长时间研究的，无非也只不过是路上的指路牌而已。但正是这些指路牌，同时也在帮助我们为将要来到的更深刻的理论研究作好准备和打造工具，并为我们照亮前进路上的坎坷。

## 更进一步的阅读

图尔明（Toulmin，1978）为我们提供了科学哲学中的强大背景。姚肖（Jasso，1988，2001，2002）的著作则为我们详述了构建社

会行为理论的程序,包括假定的选择和数学运算。这些理论对人心的鼓舞自不待言,更为基本也更为重要的是,它们对我们理解活跃的伟大理论家、假定的选择和提炼及从假定中导出寓意有很大的启示。牛顿的著作《自然哲学的数学原理》(*Mathematical Principles of Natural Philosophy*, Newton, 1686/1952)便是很好的入门读物。牛顿长期致力于基本原理的研究,后来在这个基础上提出了一种叫作假说-演绎法(hypothetico-deductive)的方法。当然,所谓最好的入门读物,是因人而异的。其他可供我们进一步阅读的著作有芒代尔(Mendel, 1866)的论文,该文是遗传学的奠基之作。还有费曼和温伯格(Feynman & Weinberg, 1987)的著作,该著作中的前言还重提了迪拉克有关反物质的预言。

——Guillermina Jasso

(沈崇麟译校)

### 参考文献


Feynman, R. P., & Weinberg, S. (1987). *Elementary particles and the laws of physics: The 1986 Dirac Memorial Lectures*. Cambridge, UK: Cambridge University Press.

Jasso, G. (1988). Principles of theoretical analysis. *Sociological Theory*, *6*, 1-20.

Jasso, G. (2001). Formal theory. In J. H. Turner (Ed.), *Handbook of sociological theory* (pp. 37-68). New York: Kluwer Academic/Plenum.

Jasso, G. (2002). Seven secrets for doing theory. In J. Berger & M. Zelditch (Eds.), *New directions in contemporary sociological theory* (pp. 317-342). Boulder, CO: Rowan & Littlefield.

Mendel, G. (1866). Versuche über Pflanzen-Hybriden [Experiments in plant hybridization]. In *Verhandlungen des naturforschenden Vereins* [Proceedings of the Naturnal History Society]. Available in both the original German and the English translation at www.mendelweb.org.

Newton, I. (1952). *Mathematical principles of natural philosophy*. Chicago: Britannica. (Original work published 1686.)

Popper, K. R. (1963). *Conjectures and refutations: The growth of scientific knowledge*. New York: Basic Books.

Toulmin, S. E. (1978). Science, philosophy of. In *Encyclopaedia Britannica* (15th ed., Vol. 16, pp. 375-393). Chicago: Britannica.


## 不对称量度(Asymmetric Measures)

关联【Association】系数测量两个变量 *X* 和 *Y* 之间的关系强度,而对定序和定量(定距和定比)变量还同时确定了关联的方向。不对称(或方向型)关联系数也假设两个变量中的某一个(如 *X*)可以被确认为自变量,而另一个变量(如 *Y*)可以被确认为因变量。

举例说明,在一个大学生调查中,变量 *X* 也许是性别(定类二分变量)、族群(定类变量)、父母的社会经济地位(定序变量)或父母的收入(定量变量)。而变量 *Y* 也许是学习专业或党派归属(定类变量)、学习成绩(定序变量)、回答人的收入(定量变量)和工会会员(答案只有"是"或"不是"两类的定类二分变量)。

### 可资利用的各种系数

大多数统计软件,如 SPSS,只给我们提供了量度层次的特定组合的不对称系数。我们先来介绍一下这些系数。

## 定类-定类表格的量度

两个定类变量的关联量度是 langmud ($\lambda$)、古德曼和克鲁斯卡尔的 tau ($\tau$)和不确定系数 $U$。所有这些量度都以误差消减比例【Proportional Reduction of Error, PRE】的定义为根据:

$$PRE = \frac{E_0 - E_1}{E_0} \quad (1)$$

式中的 $E_0$是因变量在未使用自变量 $X$ 时对 $Y$ 所做的预测或解释,而 $E_1$则是使用了 $X$ 之后对 $Y$ 的预测或解释。

$\tau$ 和 $U$ 二者都用方差的定义来计算 $E_0$ 和 $E_1$。与大家熟悉的来自方差分析【Analysis of Variance】的 eta ($\eta$)(参见下文)一样,$E_0$是 $Y$ 的总方差,而 $E_1$则是残余方差(在 $X$ 的诸类别中的方差)。$\tau$ 则用所谓的 Gini 集中系数(Gini concentration)来测量变差(Agresti, 1990, pp. 24-25),$U$ 使用了熵(entropy)。$\lambda$ 使用了一种不同的定义:分析在未知和已知 $X$ 的情形下,能在多大程度上对因变量作出预测。

所有三个系数,其变化都在 0(无消减误差)和 1(完全消减了误差)之间。即使关联清楚地存在,$\lambda$ 也可能为零或接近零。当因变量的某一类居于主导地位时便会发生这样的情况,这时 $Y$ 的这一类别有着最高的频数,因而在变量 $X$ 的每个条件上,它都成了最佳的预测项。

## 定序-定序表格的量度

用于这一类型表格的最为突出的量度是萨默斯的 $d$。萨默斯的 $d$ 被定义为因变量 $Y$ 和自变量 $X$ 之间的最小二乘回归的斜率($d_{y/x} = s_{xy}/s_x^2$),如果两个变量都是定序变量的话(Agresti, 1984, pp. 161-163)。而广义相关系数的丹尼尔斯公式(参见关联【Association】)则被用来计算 $X$ 的方差 $s_x^2$ 及 $X$ 和 $Y$ 之间的协方差 $s_{xy}$。萨默斯的 $d$ 不是 PRE 系数。相应的 PRE 系数是对称的肯德尔的 $\tau_b^2$。它可被理解为已被解释的方差;如果作为 $X$ 和 $Y$ 的因变量萨默斯的 $d$ 已经计算出来的话(等于 $d_{y/x}$ 和 $d_{x/y}$),那么它就等于几何平均数。如果两个变量都是二分的,那么萨默斯的 $d$ 便等于比例的差(Agresti, 1984, p. 161)。

## 定类-定距(或定比)表格的量度

大家所了解的已被解释的方差 $\eta^2$,或来自方差分析【Analysis of Variance, ANOVA】的 $\eta$ (eta),它们可分别应用于那些自变量为定类变量和因变量是定量变量的情形。$\eta^2$是一种 PRE 系数。

## 公式

计算上述量度的公式,可参见阿格雷斯蒂(Agresti,1990)的著作和 SPSS (2002)手册。当然也可参阅其他的统计教科书或其他软件的操作手册。

### 其他组合量度

迄今为止,我们只对配对的定类-定类对(表 1 的 1,2,5 和 6),定序-定序对(表 1 的 11)和定类-定量对(表 1 的 4 和 8)作过分析。

求其他配对,以及那些已经分析过的配对的量度的策略则如下所述:

• 使用对称量度。用丹尼尔斯的广义相关系数我们可以导出表 1 的(1,3,4,9,11 到 13,15,16)关联量度,不过那些有一个或两个多类定类变量的(表 1 的 2,5 到 8,10,14)则不在此列。

表 1 双变量量度层次的可能配对

| 自变量 $X$ | 因变量 $Y$ | | | |
|---|---|---|---|---|
| | 定类-二分 | 定类-多类 | 定序 | 定量(定距、定比) |
| 定类-二分 | 1 | 2 | 3 | 4 |
| 定类-多类 | 5 | 6 | 7 | 8 |
| 定序 | 9 | 10 | 11 | 12 |
| 定量(定距、定比) | 13 | 14 | 15 | 16 |

• 降低量度层次。将有着较高量度层次变量的量度层次降低,例如 $X$ 是定类(多类)而 $Y$ 是定序(群集 7)则可将 $Y$ 降为定类层次(群集 6),并使用 $\lambda$、$\tau$ 或 $U$ 作为其适当的量度系数。降低量度层次有可能导致信息的丢失,从而有可能掩盖存在的关联。

• 提高量度层次。将定序层次的变量作为定量变量处理。从严格的理论观点看,这一策略是不恰当的。而某些教科书(见下文)却推荐这种策略。这种策略常被用于调查研究中。在许多调查中,我经常发现定序和定量变量之间有着几乎可以说是完美的线性关系。不过在使用这种方法时,必须首先说服你自己。

• 使用 logistic 回归(logistic regression)。适当的 logistic 回归【Logistic Regression】模型被开发应用于表 1 中的许多配对(1 到 3,5 到 7, 9 到 11,13 到 15),但有些配对(4,8,12 和 16)则不在此列(Agresti, 1984, pp. 104-131)。定序因变量既可作为定量变量(Agresti, 1984, p. 114),也可作为定类变量处理。

• 使用线性回归。线性回归【Linear Regression】可用于配对 16,4 和 8。但用于配对 8 的时候,必须用虚拟变量来计算。

• 使用特殊量度。近年来统计学家们开发出了一些特殊的方法:参照单位分析(ridit analysis)(Agresti, 1984, pp. 166-168)和支配统计量(dominance statistic)$\Delta$(Agresti, 1984, pp. 166-168)可应用于集群 3 和 5,以及 11 和 15,如果因变量被作为定类变量处理的话。

• 二分变量。这样的变量可采取异乎寻常的策略。它们既可被看成定序变量,也可被看成定量变量。因此适用于作为对 11 或 1,3 和 9 的策略,而所有适用于 16 的策略也适用于 1,4 和 13。如此这般,以此类推。

• 推广。使用 logistic 或线性回归的巨大优越性在于,可对一个以上的自变量进行可靠的推广。因此,我们建议诸位使用这些高级的方法。不过 logistic 回归是解决不了关系强度问题的。与所有关联系数一样,求广义的阈值也同样是不可能的(参见关联【Association】)。

• 系数的显著性。所有已经讨论过的量度和系数都有可资利用的检验统计量。SPSS(2002)给出了它们的计算公式。

——Johann Bacher

(沈崇麟译校)

## 参考文献


Agresti, A. (1984). *Analysis of ordinal categorical data.* New York: John Wiley.

Agresti, A. (1990). *Categorical data analysis.* New York: John Wiley.

SPSS. (2002, August). Crosstabs. In *SPSS 11.0 statistical algorithms* [Online]. Available: http://www.spss.com/tech/stat/algorithms/11.0/crosstabs.pdf.

# 渐近性(Asymptotic Properties)

为了进行统计推论【Statistical Inference】,我们需要知道我们使用的估计量【Estimator】的抽样分布【Sampling Distribution】。一般来讲,精确的或有限样本的分布将是不可知的。而渐近性则通过在数据集趋于任意大(也即无限)的时候,观察估计量的性状,给我们提供了它的近似分布。

近似的结果取决于样本的大小 $n$。令 $x_n$ 为一随机变量【Random Variable】,下标 $n$ 为样本【Sample】的大小。为了定义一致性,我们需要定义概率【Probability】中收敛的记号。我们说,如果对于任何正的 $\varepsilon$,$\lim_{n\to\infty} \text{Prob}(|x_n-c|>\varepsilon)=0$,那么 $x_n$ 在概率中收敛于常数 $c$。概率中的收敛意味着变量可取的不接近 $c$ 值的可能性,将随着样本量的增大而减小。概率中的收敛通常都由 $\text{plim } x_n=c$ 来表示。这样我们就可以说,如果 $\text{plim } \hat{\theta}=\theta$,那么 $\theta$ 的估计量 $\hat{\theta}_n$ 便是一个一致估计量(consistent estimator)。大致地讲,随着样本量逐渐增大,估计量的抽样分布将堆积于真值。

虽然一致性是估计量的一个重要的性质,但为了进行统计推论,我们需要知道它的抽样分布。为了导出估计量的抽样分布,我们需要定义极限分布的记号。假设 $F(x)$ 是某个随机变量 $x$ 的累积分布函数(cdf)。此外,令 $x_n$ 为随机变量序列,下标 $n$ 为有 cdf $F_n(x)$ 的样本量。我们说,如果 $\lim_{n\to\infty}|F_n(x)-F(x)|=0$,那么 $x_n$ 在分布中收敛于 $x$。

我们将分布中的收敛记作 $x_n\xrightarrow{d}x$。用语言来表述就是,随着样本量的增大,$x_n$ 将趋近于 $x$ 的分布。这种分布中的收敛,以及中心极限定理【Central Limit Theorem】,使我们得以据此来构建一个估计量的渐近分布。

渐近分布是一种被用来近似一个随机变量的真有限样本的分布。假设 $\hat{\theta}_n$ 是参数 $\theta$ 的一致估计。那么使用中心极限定理,$\hat{\theta}_n$ 的近似分布就是:

$$\sqrt{n}(\hat{\theta}_n-\theta)\xrightarrow{d}N[0,V] \qquad (1)$$

式中的记号 $N[\ ]$ 表示 cdf 的正态分布【Normal Distribution】,而 $V$ 则是近似的协方差矩阵。这意味着 $\hat{\theta}_n$ 的分布恰如 $N[\theta,\frac{1}{n}V]$。这是一个非常令人振奋的结果。因为这表明,不论真实的潜在有限样本的分布是什么,随着样本量逐渐变大,我们都可以用正态分布来近似估计量的分布。这样我们就可以用正态分布来进行假设检验【Hypothesis Testing】,就好比我们已经知道了真的抽样分布那样。

然而必须注意的是,尽管随着样本的增大,近似程度会有所提高,但是近似分布充其量也只不过是一种近似而已;因此样本量、检验力和假设检验的其他性质也只不过是被近似而已。遗憾的是,在一般情况下,将近似分布用于统计推论时,我们无法确切地说,为了具有某种置信度,样本究竟需要多大?我们必须依靠以前的经验,自助抽样推论法【Bootstrapping】模拟或蒙特卡洛【Monte Carlo】实验来告诉我们,在任一给定情境下,逐渐逼近近似的用处究竟有多大。

——Jonathan N. Katz

(沈崇麟译校)

## 参考文献

White, H. (1984). *Asymptotic theory for econometricians*. Orlando, FL: Academic Press.

## ATLAS.ti

ATLAS.ti 是一种用于计算机辅助定性数据分析的软件(参见 CAQDAS)。因此,它允许我们对定性数据进行编码和检索。此外,它还支持可视图像和文本的分析。它也支持扎根理论【Grounded Theory】的实践,如备忘录【Memos】的生成。

——Alan Bryman

(沈崇麟译校)

## 衰减(Attenuation)

衰减是指两个不同量度之间的相关【Correlation】因量度误差而导致的降低。当然,不存在得分完全可靠的检验。不仅如此,不可靠的得分还会导致质量欠佳的标准的可预测性(predictability of the criterion)。为了求相关的估计值,如果预测子(predictor)和标准分数完全可靠【Reliability】,斯皮尔曼(Spearman,1904)建议使用如下被称作衰减的相关的公式。

$$\bar{r}_{xy} = \frac{r_{xy}}{\sqrt{r_{xx} \times r_{yy}}} \qquad (1)$$

这一公式也叫作双修正(double correction),因为预测子($x$)和标准($y$)的分数二者都被修正了。假设我们正在考察抑郁和工作满意度之间的相关性。这两个量度之间的相关($r_{xy}$)= 0.20。如果抑郁量表($r_{xx}$)分数的信度(即 Cronbach $\alpha$)是0.80,而工作满意度量表($r_{yy}$)得分的信度是 0.70。那么衰减修正相关的计算如下:

$$\bar{r}_{xy} = \frac{0.20}{\sqrt{0.80 \times 0.70}} = 0.26 \qquad (2)$$

在这个例子中,只有少量的修正。不过,假设抑郁得分的信度和工作满意度的信度分别为 0.20 和 0.30。这时衰减修正后的相关便等于 0.81。信度或样本量($N<300$)越小,衰减的修正相关就越大(Nunnally, 1978)。衰减修正的相关甚至有可能大于 1.00(Nunnally, 1978)。而在有些情况中,我们可能只想修正预测子或者标准中的某一个。例如,假设机械技能考试($x$)得分和工作完成情况($y$)得分之间相关为 0.40,而工作完成情况的信度为 0.60。如果我们只想修正标准的不可靠性,那么我们可考虑使用下面的公式:

$$\bar{r}_{xy} = \frac{r_{xy}}{\sqrt{r_{yy}}} \qquad (3)$$

在这一例子中,衰减的修正相关等于 0.51。尽管很少有研究者只想修正考试分数,但实际上这种做法也是可行的。而在标准的信度非常好或未知的时候,就有可能发生这样的修正(Muchinsky, 1996)。这时,我们可能需要使用下面这样的公式:

$$\bar{r}_{xy} = \frac{r_{xy}}{\sqrt{r_{xx}}} \qquad (4)$$

马金斯基(Muchinsky,1966)对衰减修正的原理作了总结。首先,衰减的修正不会提高考试分数的可预测性(Nunnally, 1978)。不仅如此,修正的相关既不能做统计显著【Statistical Significance】检验(Magnusson, 1967),也不能进行修正的和未修正的效度【Validity】系数之间的相互比较(American

Educational Research Association, 1985)。衰减的修正不能用于效度系数的均差型分析,如元分析【Meta-Analysis】。因为元分析使用了不同的信度估计(例如,内部一致性【internal consistencies】,测验-再测)(Muchinsky, 1996)。

——N. Clayton Silver

(沈崇麟译校)

## 参考文献

American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. (1985). *Standards for educational and psychological tests*. Washington, DC: American Psychological Association.

Magnusson, D. (1967). *Test theory*. Reading, MA: Addison-Wesley.

Muchinsky, P. M. (1996). The correction for attenuation. *Educational and Psychological Measurement*, *56*, 63-75.

Nunnally, J.C. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.

Spearman, C. (1904). The proof and measurement of association between two things. *American Journal of Psychology*, *15*, 72-101.

# 态度测量(Attitude Measurement)

态度是依据某种从赞许到不赞许排列的维度,对某一对象(人、地方或事情)作出的评价(Weiss, 2002)。正因如此,态度都涉及情感(情绪)和认知两个方面。因为它是人的内在状态,所以它可以用定量的心理量表,通过回答人的自我报告来确定。心理量表可以有几种不同的形式(Judd, Smith, & Kidder, 1991)。

## 瑟斯通量表

瑟斯通量表(Thurstone & Chave, 1929)由一个对态度对象(attitude object)的赞许性变化的题项清单组成。它的基本思想是个体将只认同那些与他们所持态度位置相近的题项。例如斯佩克特等(Spector, Van Katwyk, Brannick, & Chen, 1997)创建了一套工作态度量表,它的题项涵盖了极端负面(我恨我的工作)到极端正面(我的工作是我生活中最令我感到快乐的事情)及位于二者之间的比较正面的题项(我喜欢我的工作)和比较负面的题项(我不太喜欢我的工作)。

瑟斯通量表的构建始于将一组判断写成一组题项,然后将写就的题项从极不赞许到极为赞许分类排序。之后再对已经排序的题项进行编号,并计算跨各种不同判断(题项)的平均分。在使用这种量表时,要求回答人读出提交给他们的随机排列的题项,并对每一个题项明确指出同意还是不同意。态度得分是回答人标明的题项量表值的平均值。

## 李克特(累加评级)量表

李克特或累加评级量表【Likert or Summated Rating Scale】(Likert, 1932)是由一系列对态度客体的赞许或不赞许的陈述组成的。回答人要对给予他们的有着多种选择题项的答案作出选择。通常这些多项选择沿着某一连续统变化,变化通常都是从同意到不同意。回答人要对每一问题表明自己的同意(或不同意)程度。这种量表的设计基于这样一种假定:赞许和不赞许题项是互为镜像的,这样某一个体如果同意某一组题项,即意味着他

将不同意另一组题项。这一假定与瑟斯通量表大相径庭,瑟斯通量表认为人们只同意赞许性或不赞许性与自己所持态度位置相近的那些题项。斯佩克特等人(Spector et al.,1997)证明李克特量表的假定并非总是正确的,而在用诸如这样的题项作因子分析时,有可能会产生人为制造的结果。

李克特量表已经成为在态度评定中使用最为普遍的量表,其主要原因无疑有二。首先它比较容易构建,在获得题项的判别等级时,并不需要作最初的尺度研究。其次,这种方法的量表已经产生了不少有较高信度、对态度评定很有用处的量表。

## 语义差异量表

语义差异量表【Semantic Differential Scale】是由奥斯古德、苏吉和坦嫩鲍姆(Osgood, Suci, & Tannenbaum,1957)开发的,用来对对象态度的不同方面进行评估。该量表的主干是态度对象的姓名或描述(例如,美国高等法院)。研究者提供了一系列双极形容词,在两极之间提供了一系列答案选择(排列有序的横杠)。回答人则被要求在那个最能反映他们情感的备选答案(相应的横杠)上打钩,回答人的选择(打钩的横杠)越接近形容词,则表示对该形容词的感情越强烈(参见表1)。

**表1 语义差异例示**

| 美国高等法院 | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 好 | —— | —— | —— | —— | —— | —— | —— | 坏 |
| 公正 | —— | —— | —— | —— | —— | —— | —— | 不公正 |
| 弱 | —— | —— | —— | —— | —— | —— | —— | 强 |
| 消极 | —— | —— | —— | —— | —— | —— | —— | 积极 |
| 高兴 | —— | —— | —— | —— | —— | —— | —— | 不高兴 |
| 不重要 | —— | —— | —— | —— | —— | —— | —— | 重要 |

语义差异题项通常把态度归并成三个潜在的维度:评价、效价(potency)和积极性(activity)。评价表明了该人所持赞许或不赞许的程度,也就是说不论其喜欢还是不喜欢该对象。效价指对象的强度,表明一个人把态度对象视为强有力或否的程度,不论其是否对它有好感。积极性是指把态度对象视为主动或被动的程度。

语义差异常用于同时评价对几个对象的态度,而不是只为一个对象开发一个态度量表。我们可以像跨题项得分那样,一次只包含一个评价两极化形容词,产生每一态度客体的一个单独的得分。对于回答人来讲,这样的方法可能更为容易和有效率。因为这种量表,比瑟斯通量表或李克特量表更容易读,且可将同样的题项用于每一态度对象,从而更容易作比较。

## 哥特曼量表

哥特曼量表【Guttman Scale】(Guttman, 1944)是由一组呈现从不赞许到赞许变化的等级题项组成的。它要求回答人对每一题项表明同意还是不同意。它假定持赞许态度的个人将会同意那些位于相应层次或较低层次的题项。因此一个持有些赞许态度的人,会同意有些赞许的题项,而不会同意

极为赞许的题项,而一个持极为赞许态度的人,将会同时同意有些和极为赞许二者。哥特曼量表不像其他量表那样常为研究者使用,其原因主要有两个:首先较之使用更为普遍的李克特量表和语义差异量表,哥特曼量表比较难以开发。其次,期望的答案模式并不总是能够出现的。这是因为那些持极端态度的人常常不会同意那些有些赞许的题项,或这些题项反映的并非一个单一的态度维度。

——Paul E. Spector

(沈崇麟译校)

## 参考文献


Guttman, L. (1944). A basis for scaling quantitative data. *American Sociological Review*, *9*, 139-150.

Judd, C. M., Smith, E. R., & Kidder, L. H. (1991). *Research methods in social relations* (6th ed.). Fort Worth, TX: Harcourt Brace Jovanovich.

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, *140*, 55.

Osgood, C. E., Suci, C. J., & Tannenbaum, P. H. (1957). *The measurement of meaning*. Urbana: University of Illinois Press.

Spector, P.E., Van Katwyk, P.T., Brannick, M.T., & Chen, P.Y. (1997). When two factors don't reflect two constructs: How item characteristics can produce artifactual factors. *Journal of Management*, *23*, 659-677.

Thurstone, L. L., & Chave, E. J, (1929). *The measurement of attitude*. Chicago: University of Chicago Press.

Weiss, H. M. (2002). Deconstructing job satisfaction: Separating evaluations, beliefs and affective experiences. *Human Resources Management Review*, *12*, 173-194.


# 属性(Attribute)

属性可以定义为一个受访者、研究对象或个案所具有或不具有的特征。二分属性变量有两个值:这一属性存在或不存在。例如,在普查中的成年人可能具有已婚这一属性(取值=1)或没有这一属性(取值=0)。心理学或者市场研究则将属性定义为可计量的项目。例如,决策研究中的属性可能指的是与决策者相关的目标或标准。举例来说,在选新车的时候,购买者可能会面临三种可选择的(属性),如燃油效率、费用和安全性,因而必须对三者加以权衡。

——Michael S. Lewis-Beck

* 也可参见二分变量【Binary, Dichotomous Variables】。

(杨可译 沈崇麟校)

# 损耗(Attrition)

损耗这个术语用来描述在定量调查的资料收集过程中,由于无回答【Nonresponse】及调查对象不再符合参与资格而导致样本规模不断减小的过程,这个术语也用来表示这个过程的量化结果。

在定群【Panel】调查中,由于无回答的累积效应和参与资格适当性的变化,每一波参与回答的样本单位数量都在持续减少。由于拒答的非随机性,这样的损耗可能带来对调查估计的难以测量的偏差。因此,研究

者一直都在非常努力地减少样本损耗程度。损耗程度也会受到资料收集的期数(waves)、每期资料收集的间隔时长、资料收集的工作量等因素的影响。为了限制损耗的不利影响,研究者会在分析阶段使用加权【Weighting】及其他调整方法。

尽管损耗现象在定群调查中最为明显,但实际上损耗在任何具有多阶段资料收集过程的研究中都会产生影响。许多横断面【Cross-Sectional】调查都会使用多个数据收集工具,例如,被调查者在接受访谈之后,进行的自我完成【Self-Completion】(“留答问卷”)问卷【Questionnaire】,或在访谈后进行一次体检。在这样的情况中就会产生不同研究阶段之间的损耗,比如并非所有被访谈者都会完成并交回自我完成问卷。即使研究中只涉及单一的资料收集工具,调查研究过程也可能有多个阶段。例如,如果要开展一项针对在校小学生的研究,必须首先得到被选入样本范围的学校的合作,这些学校能够提供可被选为样本的学生名单。接下来在接触学生本人之前还必须得到家长的知情同意。因此,这项研究中至少存在三个可能使样本量减小的阶段:学校的无回答、家长的无回答和学生本人的无回答。这种情况也会被看作样本损耗。

在解决损耗问题的过程中,无论是通过从源头减少损耗,还是在分析阶段进行调整,都需要认识到不同要素的子进程,并予以仔细的考虑。在产生损耗的原因方面,符合调查资格的被访者无回答,与曾符合资格的被访者不再符合资格,这二者具有重要的概念差异。后一种情况如果能够被恰当识别,不一定会对分析产生不利影响。这两种损耗的主要情况会分别产生不同的从属情况:无回答的原因可能是无法联系到被访者(“无法联系”),被访者无能力回答(例如由于语言、疾病等原因),或者被访者不愿意回答(“拒绝”)。被访者合适资格的变化,原因可能是死亡、居住地点改变、身份地位变化,等等。不同情况对于损耗的减少和调整有不同的含义。许多定群调查花费大量资源来长期“追踪”样本研究者,以便减少无法取得联系的比率。研究者认为,参加某一次调查的经历(花费的时间、兴趣、敏感度、认知需求等)是决定后续调查拒访率的重要因素;因此,研究者力图使被调查者最大限度地感受到积极合作的获益,最小限度地感知到参与调查的不利和障碍。

——Peter Lynn

(郑晓娟译　沈崇麟校)

## 参考文献


Laurie, H., Smith, R., & Scott, L. (1999). Strategies for reducing nonresponse in a longitudinal panel survey. *Journal of Official Statistics*, *15*(2), 269-282.

Lepkowski, J. M., & Couper, M. P. (2002). Nonresponse in the second wave of longitudinal household surveys. In R. M. Groves, D. A. Dillman, J. L. Eltinge, & R. J. A. Little (Eds.), *Survey nonresponse* (pp. 259-272). New York: John Wiley.

Murphy, M. (1990). Minimising attrition in longitudinal studies: Means or end? In D. Magnusson & L. R. Bergman (Eds.), *Data quality in longitudinal research* (pp. 148-156). Cambridge, UK: Cambridge University Press.

## 审核(Auditing)

审核是一种确立定性研究的结论可靠性的方法。可靠性与可信性标准【Trustworthiness Criteria】有关。审核要求保留定性研究项目各个阶段的完整记录,并将其提交给同行,以便他们评定研究是否遵循了正当的程序,包括理论推理的稳健性以及研究实践是否合乎伦理。

——Alan Bryman

(杨可译　沈崇麟校)

## 真实性标准(Authenticity Criteria)

真实性标准用来评判那些以自然的、非实证研究范式来进行的调查(包括研究、评估和政策分析)的质量,常使用该研究范式的包括批判理论【Critical Theory】、参与调查、建构主义【Constructivism】等。我们已经建立了严格的实证性标准:内在效度【Internal Validity】、外在效度【External Validity】、信度【Reliability】和客观性【Objectivity】。之前,我们试图将评判研究质量的,那些类似的但非实证的方式称为可信性标准【Trustworthiness Criteria】(Guba, 1981)。与实证性标准并行,我们对可信性标准也给出了四个评价维度,分别叫作可信性、可转移性、依赖性和验证性。但是,与严格实证性标准一一对应的这些特性,在对于非传统调查的适用性方面遭到了质疑。此外,标准性维度主要是方法论意义的,所关注的议题包括权力、伦理、声音、调查通道【Access】、表征或其他后现代主义【Postmodernism】所提出的议题。

真实性标准是伴随着非传统研究范式的提出而发展起来的(Guba & Lincoln, 1989)。虽然这套标准还是不完整和比较初步的,但它仍不失为一个合理的指导原则。

目前一共有五个真实性准则,它们都有特定的先决条件(initial conditions),包括:被调查者需要取自所有处于风险中的群体,必须将完全的知情同意【Informed Consent】程序置于适当的地方,培育出相互关怀和信任的情感关系,调查程序对所有的被调查者及观众都是透明的,调查的每个步骤中都会建立起被调查者和研究者的合作关系,包括对实施调查的规则的完全同意和信息的充分共享。调查报告要保证向每个被调查者和观众开放。最后,应建立相应的申诉机制,以应对冲突或意见不一致情况的发生。

这五个标准的定义及为了建立这些标准所推荐的程序包括如下内容:

- 公平性,用来描述所有有竞争关系的实体建构,以及其隐含的价值结构已被存取、显露、解构,并在调查产物中予以考虑,也就是自然重构的程度。公平性标准需要遵循严格的程序。被调查者和研究者的所有先前建构,都将在相似的基本权利下被获取、比较和对比;前后的调查建构需要进行比较和对比;被调查者和研究者需要就资料收集的方法、使用的方法、要做的解释、撰写报告采用的模式、将要做出的推荐和建议(采取)的行动等问题进行协商;需要收集被调查者和研究者关于所经历的变化的自省式陈述(证言);使用长时间的参与和持续观察;采用个体和群体成员的检查/效验;提供丰富的背景性描述;使用同期群中的情况汇报者和审计员(意为维持审查路径);以及最后,评估被调查者感知到的被赋权程度。

• 实体论标准，被定义为个人被调查者（和研究者）前期的建构被改进、完善、扩充、详细化，以便所有获得该研究的相关方能够获取更多信息，并在该研究的应用方面更加得心应手。有用的程序包括：对被调查者和研究者前期立场的说明；对被研究者前期和后期个人解释的比较；对被调查者和研究者关于自身成长的自省式陈述和选中的被调查者关于自身变化性解释的证言；建立揭示变化的审查路径。

• 教育性标准，被定义为个体被调查者（和研究者）具有的，使他们自身的利益群体之外的人，对建构的了解、欣赏和容忍的增强程度。有用的程序包括：研究者使用同期群中的汇报者和审查员，对于研究者和被调查者对他人看法的评价进行比较，被调查者和研究者关于他们对他人观点的理解的自省式陈述，被调查者证言和审计路径的保持。

• 接触反应标准，被定义为行动对（阐明议题关注点，转移、消除或改善问题，和/或形塑价值）被调查过程触发和促进的程度。这个过程中的知识本身对于被调查者带进研究过程的要阐明的建构、问题、关注点和议题是不够的；有目的的行动也需要加以描绘。有用的程序包括发展出一套共同观点（致力于达成可能的一致意见，或者当一致的意见不可能达成时，对冲突的价值进行解释），包括在行动中赋予责任和权威，被调查者—研究者合作的计划，对最终报告的可获得性以及实际应用的证据。参与者证言，至少实际解决部分调查关注点，对结果的持续价值进行长期系统性评估也是很有帮助的技术。

• 策略标准，被定义为研究所暗含或明确提出的为参与者赋予行动权利的程度。有用的程序包括对收集资料及其解释和报告的协商；保密【Confidentiality】；知情同意形式的使用；成员检查/效验；对权力的事先协定。这个标准被满足的最好指标就是参与者证言。跟进活动的存在则是一个灵敏的指标。最后，参与者所感知的赋权程度也是至关重要的。

真实性标准的发展现状还有许多的改进空间。或许目前这方面最重要的成就在于它的存在本身，它揭示了我们有可能在传统质量评估的“盒子以外思考”。研究者非常有必要去思考他或她所选择遵循的范式，并决定选出合适的质量标准。对非传统研究范式采用实证性标准，或反过来对传统研究采用真实性标准都只会造成困扰和冲突。

——Egon G. Guba

（郑晓娟译　沈崇麟校）

## 参考文献

Guba, E. G. (1981). Criteria for assessing the trustworthiness of naturalistic inquiries. *Educational Communication and Technology Journal*, *29*, 75-92.

Guba, E. G., & Lincoln, Y. S. (1989). *Fourth generation evaluation*. Newbury Park, CA: Sage.

# 自传（Autobiography）

自传是指对自身生活的讲述和记录。与传记（对他人生活的研究和记录）类似，自传越来越成为一种研究社会生活的资源和方法。自传研究是社会科学中更广义的传记研究转向的一部分，其特征是强调个人叙事和生活史方法【Life History Method】。

## 自传作品的历史

无论在宽泛意义上，还是在社会科学领域内，自传和生活史作品都有一本很长的家谱。肯·普吕默(Plummer, 2001)追溯了“自传社会”的兴起和已经发生的几次转型。他记录了伴随从口述向书写传统的变化，对生活的讲述变成了用文本书写的方式。他发掘了自传形式可能的起源和个体“声音”的兴起。他还研究了个体的生活故事如何成为探究共享生活和经验的组成部分。利兹·斯坦利(Stanley, 1992)追溯了社会科学(尤其是社会学领域)和女权主义研究领域中自传的作用。她仔细研究了自传著作的方法论问题，强调女权主义实践与自传实践的关系。女权主义运动对当代社会科学中自传和个人叙事的发展产生了特殊的影响，她们既把自传当作了一种“发声”方式，也看作一种强调反身性和个体知识的路径。在更广泛的意义上，自传和生活史作品能够为那些会被隐藏或消除的声音，如边缘或被压迫的社会群体，创造文本和推论的空间。

## 自传的资源和叙事

社会科学家会将自传资料视为一种资源。通过个人记述和生活故事可以揭示出生活和经历的普遍特征。书面自传反映了关于生活、场景、组织、文化、事件或某一特定时刻的内容，因而自传本身就可以作为一套可挖掘和分析“生活”的丰富资料集。在这个意义上，生活故事可以理解为对社会生活的真实记录。同时，自传作品还是经过创作，并通过文本形式表达的。因此，可以将自传本身当作一个研究事件——是生活被记忆、重构和书写的过程。对于以关键事件、地点、人物和时间为标志的生活，存在共同的记忆习惯和讲述(重述)“方式”。因而自传文本也存在能够被系统发掘的特定体裁和叙述形式。它们是一套建立在标准文学和文本传统上的故事。这些故事的结构和形式能够通过诸如叙事分析【Narrative Analysis】那样的方法来分析。

## 自传和方法(论)

自传作品所使用的“方法”也可以用来收集社会科学资料。尽管自传本身并不是一种社会科学研究方法，但它确实使用了一系列的经验材料或“生活文献”(Plummer, 2001)。这些资料包括信件、日记【Diary】和其他形式的文献资料，手工制品和传记性物品，电影、视频、照片等视觉媒体资料。尽管这些资料源有巨大的分析潜能，但在社会科学研究中它们相对而言并没有得到充分利用。因此，对于“生活”被记忆、援引、建构和再生产的关注，能够让我们对于认定社会研究资料的范围采取一种更加兼容并包的路径。

基于对女性主义社会科学的关注，斯坦利(Stanley, 1992)指出，自传(和传记)研究工作可以概念化为一种特定方法论的一部分。其中包含对反身性【Reflexivity】的关注；拒绝传统的二分法，如客观性—主观性、自我—他人、公—私；强调研究和理论化的经验。因此，社会科学研究中的(自传)传记研究路径可能意味着对研究过程、研究关系和研究者自身具有更强的自我意识，以及将生活经历【Lived Experience】和个体认知作为社会科学学识组成部分的价值给予明确赞赏。

## 书写和研究自我

对研究过程的个人叙事本身就是一种自传实践。社会研究者，同时也是他人生

活的记录者，他或她在书写他人自传的同时，也在进行自传的书写。特别是在定性研究【Qualitative Research】中的那些工作，它们以各种各样的自传文本反映和书写了自身。长时间以来，田野笔记【Fieldnotes】和研究日志（日记）一直被用作记录与社会研究相关联的情绪、经历和认同努力（identity work）。告解故事（confessional tale）这一体裁——揭示了个人田野工作的某些方面——在被认可的自传形式内，允许自我表露失检的言行。尽管这样的记述通常都被排除在研究“资料”以外，但将自传与分析混合起来还是可能的。例如，在《制定自我》（*Crafting Selves*，Kondo，1990）一书中，关于日本工作场所的“故事”就是一种挑战研究者和被研究者、自我和他人等二分观念的，多层次文本呈现。孔多的生活故事形成了研究表现的组成部分。她将对自身和她的传记作品的揭示整合进了研究过程和研究成果。

自我还能够成为外部调查的主题。因此，社会研究的方法可以明确用来承接自传著作。例如，研究者使用民族志方法来发掘研究者自身的个体经验和生活事件。这些经历包括对疾病、情感关系和丧失亲人的自传分析。这些个体的自传叙事可被看作更宽泛的自我民族志【Autoethnography】的一种类型（Reed-Danahay，2001）。

## 评估

由于自传可能使用许多不同的具体方法，我们很难将其作为一种社会研究方法加以评估。与所有处理生活故事的研究类似，自传研究应当能够反映被访者的声音和被讲述者的故事。讲述者的声音是有差异和分层的，在对待他们的生活故事时，我们固然不应对强势群体赋予特权，我们同样也不应为弱势群体代言。

对于自传式社会研究的一种批评，始终指向自传式社会研究的特定形式，认为它们可能具有美化自我或自我放纵的倾向。尽管社会研究和自传实践可以进行区分，但我们也应承认它们之间的互文性（intert-extuality）。自我会不由自主地与社会研究过程产生关联，对这一点我们应当承认和理解。总之，个体经验和自传故事能够成为富有洞见的分析和创新性社会科学的资源。

——Amanda Coffey

（郑晓娟译　沈崇麟校）

## 参考文献


Kondo, D. K. (1990). *Crafting selves: Power, gender and discourses of identity in a Japanese workplace.* Chicago: University of Chicago Press.

Plummer, K. (2001). *Documents of life 2.* London: Sage.

Reed-Danahay, D. (2001). Autobiography, intimacy and ethnography. In P. Atkinson, A. Coffey, S. Delamont, J. Lofland, & L. Lofland (Eds.), *Handbook of ethnography* (pp. 405-425). London: Sage.

Stanley, L. (1992). *The auto/biographical I: Theory and practice of feminist auto/biography.* Manchester, UK: Manchester University Press.


# 自相关（Autocorrelation）

参见序列相关【Serial Correlation】。

## 自传式民族志(Autoethnography)

自传式民族志的调查者借助自身经验，以获得对其所属的更广的文化或亚文化的深刻理解。自传式民族志将自传式的写作和民族志的研究方法相结合,将个人和文化两种透镜整合在一起,进而进行创造性的写作。

民族志【Ethnography】首先是作为一种研究和理解他者【Other, The】的方法出现的。

异文化他者的魅力曾吸引欧洲学者对非洲、亚洲、南太平洋诸岛以及美洲进行研究。在美国,对受过大学教育的白种美国人来说,他者就是黑人、美洲印第安人、新移民以及市中心的贫民。就民族志学者作为参与式观察者报告其个人经验而言,主要是在方法论方面说明他们如何收集资料或他们保持独立的程度等。

与之相反,自传式民族志的调查则会提出如下的问题:我自身在本文化中的经验是如何为我提供对于我的文化、境况、事件,与/或生活方式的深刻理解的?自传式民族志提出了一个问题:一个民族志学者是如何得以在没有负担、不假装独立的情况下研究自身的文化(Patton, 2002)?如果没有一个他者作为研究的焦点,但我想研究我自己的群体、自己的社区、自己的组织、像我一样的人的生活方式、我常常遇见的那些人或者我自己的文化经验,又该怎么办?

据称大卫·海雅诺(Hayano,1979)发明了自传式民族志(autoethnography)一词来描述人类学家对自身文化的研究。

自传式民族志通常以第一人称进行写作,但也可以是小故事、诗歌、小说、摄影、个人随笔、日志以及社会科学的文章。尽管自传式民族志是运用自身经验来获得对其所属的更广的文化或亚文化的深刻理解,但各种自传式民族志在何种程度上将作者自身作为分析的焦点,作者多大程度上在前台保持自己的社会科学研究者身份,在何种程度上(至少是明确地)运用自己对文化的敏感理解来指导其分析,以及写作究竟有多个人化都有很大的差异。不过,就其核心而言,自传式民族志与民族志的区别在于对自我意识的认识,报告个人的自我经验,以及将反思作为主要的数据来源。卡罗琳·埃利斯(Ellis & Bochner, 2000)曾对这个过程作了如下描述:

> 我从我的个人生活开始。我关照我的身体感受、思想和情绪。我采用我所说的系统的社会学反思和情感回忆,试图理解我经历的体验。接下来我把我的经验当成故事来写。通过探索这一独特的人生,我希望能了解一种生活的方式。(p. 737)

### 界限模糊的查询

自传式民族志模糊了社会科学和文学写作之间的界限,对此很多社会科学的学者都不赞成。但劳雷尔·理查森(Richardson,2000)却正是在这里看到了艺术、文学和社会科学的结合,看到了调查中创造性和批判性两方面的融通。她认为,这些各式各样的新方法和研究重点有一个共同点,就是"它们都是由创造性的分析实践生产出来的",由此,她将"这一类的民族志"称为"创造性的分析实践的民族志"(p.929)。还有一些其他的术语被用于描述这种界限模糊的、引人共鸣的定性调查形式——例如人类学诗学、民族志诗学、民族志回忆录、民族传记、共鸣式叙事、实验民族

志、本土民族志、文学民族志、本地人民族志、个人民族志、后现代民族志、反思民族志、自我民族志、社会民族志和社会诗学，等等(Patton, 2002, p.85)。

不过，人们怎么判断这类非传统的、鼓励个人化与创作式的民族志写作的社会科学方法的质量呢？除了通常要求的实质性的贡献和方法论的精确以外，自传式民族志以及其他的引人共鸣的研究形式还要加上审美特征、反思【Reflexivity】(对个人观点及其影响因素的自觉)、对读者的影响力，以及表达的真实性等标准。沿着这一脉络，埃利奥特・艾斯纳(Eisner, 1997) 提出，“在定性研究方法的前沿领域”，艺术性的定性社会科学的贡献不仅可以由其提供的结论来评价，同样可以通过“作品提出的问题的数量和质量”来评价(p.268)。自传式民族志也可以采用非小说类创作的表现形式。例如，社会学家迈克尔・奎因・巴顿(Patton, 1999)曾描写了与儿子一起度过的大峡谷十日徒步行，在这次旅行中他们探索了在现代社会里所谓达到法定年龄或者进入成年究竟意味着什么。为了让这一研究像个故事，而且使分散的互动保持连贯，他把几天里发生的谈话都纳入了一场晚间的对话里，他改变了一些对话的顺序以突出情节线索，还把自己的情感、癖好、怀疑、弱点和疑虑都当作调查资料透露出来。

自传式民族志把民族志和个人经历联系在一起，而后者属于广义的社会科学传记法【Biographical Methods】里特殊的一种，试图“将宏观和微观层面的分析联系起来，……(而且)为个人与社会的关联提供一个复杂的阐释性的程序”(Chamberlayne, Bornat, & Wengraf, 2000, pp. 2-3)。尽管充满了争议，但社会科学和艺术之间模糊的边界还是创造了调查的新形式，开启了富于创造且引人共鸣的报告发现的新方式。

——Michael Quinn Patton

(郑晓娟译　沈崇麟校)

## 参考文献


Chamberlayne, P., Bornat, J., & Wengraf, T. (2000). *The turn to biographical methods in social science*. London: Routledge.

Eisner, E. W. (1997). The new frontier in qualitative research methodology. *Qualitative Inquiry*, *3*, 259-273.

Ellis, C., & Bochner, A. P. (2000). Autoethnography, personal narrative, reflexivity: Researcher as subject. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (2nd ed., pp. 733-768). Thousand Oaks, CA: Sage.

Hayano, D. M. (1979). Autoethnography: Paradigms, problems, and prospects. *Human Organization*, *38*, 113-120.

Patton, M. Q. (1999). *Grand Canyon celebration: A father-sonjourney of discovery*. Amherst, NY: Prometheus.

Patton, M. Q. (2002). *Qualitative research and evaluation methods*. Thousand Oaks, CA: Sage.

Richardson, L. (2000). Writing: A method of inquiry. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (2nd ed., pp. 923-948). Thousand Oaks, CA: Sage.


# 自回归(Autoregression)

参见序列相关【Serial Correlation】。

# 平均数(Average)

平均数这一术语用来表示一个变量的典型值,特别是那些表明分布的中点、均值或中位数的值。一般地讲,它都指的是算术平均值。算术平均值是一给定的变量值之和除以个案数。中位数是值的分布的中点。正因为这样,恰好有一半的值大于中位数,而有一半的值小于中位数。至于我们究竟应该选择哪一种平均数,则取决于不同场合中,这两个量度的特定优点和缺点。这个问题将在后面的集中趋势量度【Measures of Central Tendency】词条中进一步讨论。

## 例子

尽管大多数学生对数学可能不甚了了,但却能计算出自己的成绩的平均分,因为课程的期末成绩通常都是平时所有作业(在本例中共有三个作业)成绩的算术均值。表1是假设的学生学习的某门共有三个作业的课程的作业成绩,每个学生的学习成绩的均值和中位数,以及跨不同作业的成绩的分布。它显示不同的分布,学生作业成绩高于或低于平均分的不同方式,均值和中位数有不同的关系。Dudley 三次作业成绩都很好,因此期末成绩不仅同样也很好,且均值和中位数很接近。而 Jenny 的期末成绩虽然同样也很好,但三次作业的成绩有着较大的变差,导致中位数明显高于均值。此外,Jenny 第一次作业非常出色,高出均值 17 分,高出中位数 19 分。而所有学生的第二门作业都同样好,都在 61 分以上。第一次作业的难度显然高于其他两次,因此这次作业的分数应该有更高的加权【Weighting】。这样 Jenny 的期末成绩就会高于 Dudley。该表也是一个中位数如何围绕均值变化的简单例子。当然,究竟怎样做才会使期末成绩最为公平合理,这是一个教育界长期争论不休的问题。

**表1 九个学生三次作业成绩和作业平均成绩(人为构造的数据)**

| | 第一次作业 | 第二次作业 | 第三次作业 | 课程作业平均分 | |
|---|---|---|---|---|---|
| | | | | 均 值 | 中位数 |
| Adam | 52 | 62 | 63 | 59 | 62 |
| Akiko | 51 | 63 | 55 | 56.3 | 55 |
| Chris | 55 | 65 | 57 | 59 | 57 |
| Dudley | 63 | 62 | 64 | 63 | 62 |
| Jenny | 71 | 66 | 52 | 63 | 66 |
| Maria | 48 | 64 | 65 | 59 | 64 |
| Robin | 41 | 61 | 47 | 49.7 | 47 |
| Surinder | 47 | 67 | 60 | 58 | 60 |
| Tim | 58 | 66 | 42 | 55.3 | 58 |
| 均值 | 54 | 64 | 56 | 58 | 59 |
| 中位数 | 52 | 64 | 57 | 59 | 60 |

## 历史发展

将重复观察组合起来使用,以检查每次观察的精确性的做法可追溯到 16 世纪蒂乔・布拉厄进行的天文观察。而这种将算术均值作为一种验证形式的应用在天文学中沿用至今,但在 18 世纪和 19 世纪平均数的发展却是为了加总和估计,它与哲学上围绕自然和人类地位的争论密切相关。凯特勒这位社会科学数据分析的先驱,发明了"平均人"这一概念。他将基于这一概念构建的量度用作理想的参照点来对其他的男人和女人作比较。将自己的成绩归功于凯特勒的高尔顿,则更多地关心平均数的偏差,强调"遗传的天才"和优生学,并在此基础上导出了他对回归的理解。自此之后,"平均数"的(统计)品质才能更多地与平常而非完美相连。

## 平均数和低收入量度

我们常用平均数来概括收入和收入的分布。有些国家,特别是那些没有官方贫困线的国家,常将平均资源的分数(fractions of average resources)作为事实上的贫困线。至于用什么分数、什么平均数和什么资源量度(例如收入、消费),则随来源和国家的不同,以及那些实施量度的人的不同而不同。例如在英国,《低于平均收入的家庭》(*Households Below Average Income*)年报系列则将平均收入的 50%作为低收入定义,该定义已经成为准官方的定义。然而系列报告也引用其他的均值分数和中位数分数。作为比较,我们来看一下欧盟的贫困线。欧盟将消费的 60%作为贫富的分界,即贫困线。英国和欧盟二者都根据家庭户的大小通过一个等价的尺度对收入进行加权。诚如我们已经指出的那样,假如收入分布的峰态偏右,那么它的均值将高于中位数。因此人们一般都认为中位数较之均值,更好地概括了分布的中心。然而,一旦我们使用了平均数的分数,问题的性质或多或少,在概念和实践两个方面都有所变化。在概念上看,低收入量度试图抓住一个表明人们脱离了参与期望的,或为人们普遍认同的生活标准和实践的点。至于从中位数和均值的角度看,这一量度是否是它的最佳表述,则在一定程度上取决于贫与富的关系是怎样构建的。无独有偶,从实践的角度看,一个低收入量度旨在将收入分布的低端数量化。从均值和中位数角度看,这一量度是否达到了最佳的上限,与均值和中位数是否是收入分布的最佳概括是两个不同的问题。实际上,在收入分布中,均值的 50%和中位数的 60%落入非常相近的点。所以这两个量度是人们最为常用的,因为它们是高效互换的。

图 1 给出了一个收入分布的实例。该图显示了 1980 年代后期,美国家庭户的总收入。该图以家庭户的频数标绘了收入的直方图。绘图时数据作了截删,即将那些极端的(最高收入)数据删除掉了。最远(右)端的双线是分布的均值(每月 2 447 美元),它左边的是实线,是分布的中位数(每月2 012美元)。位于最左边峰尖上的是众数,用一条虚线表示。该图清楚显示大部分的家庭户都位于均值的左边。50%的均值和 60%的中位数是非常接近的(分别为 1 261 和 1 223 美元)。因此,它们将去掉分布的第一个峰态区域的大部分,这一部分可被视为根据不平等量度要构建的低收入家庭户。

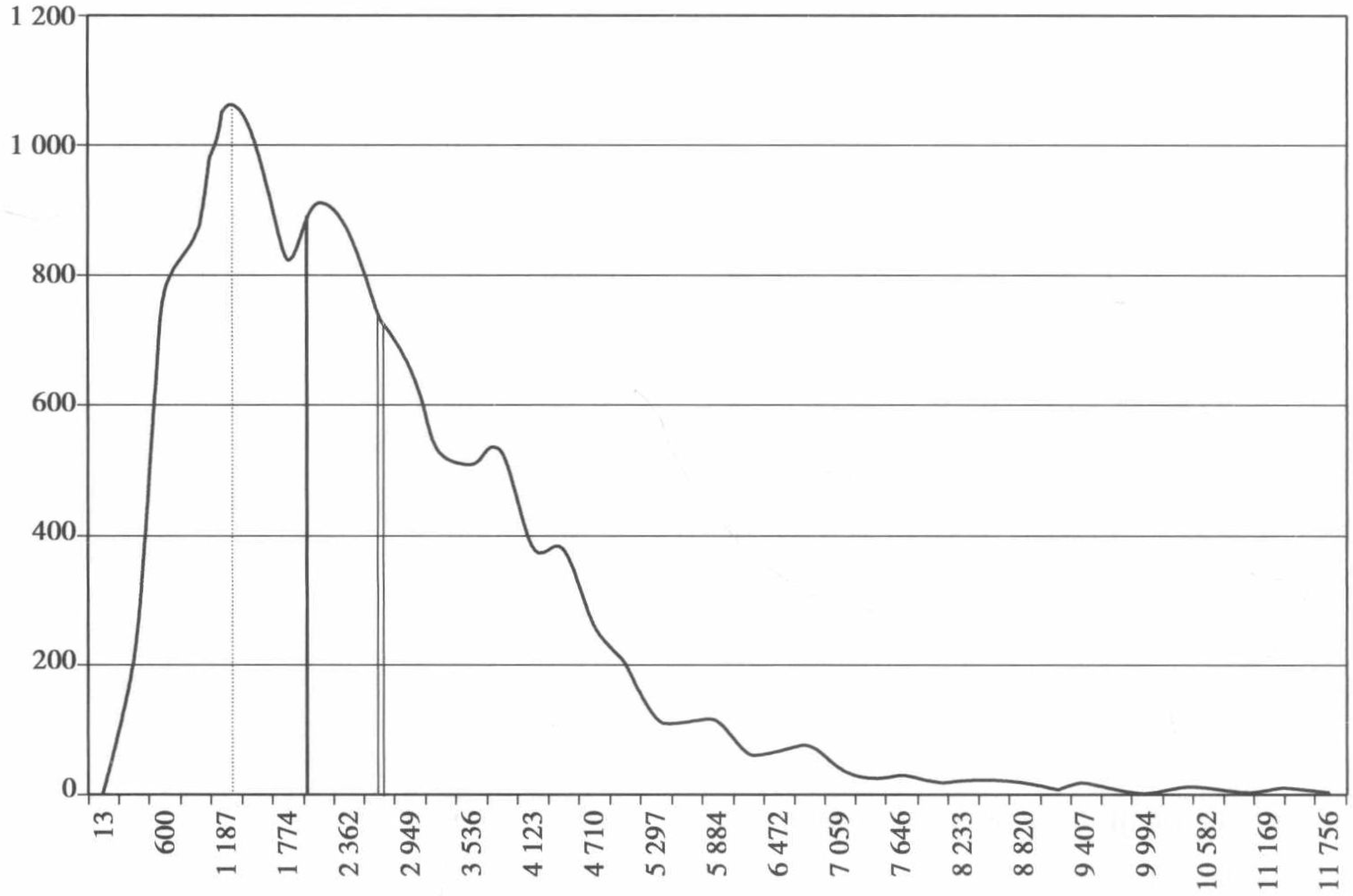



**图 1　显示众数、中位数和均值的典型收入分布图**

（资料来自 1987 年美国家庭户收入调查；数据已作截删）

——Lucinda Platt

（沈崇麟译校）

＊也可参见均值【Mean】、集中趋势量度【Measures of Central Tendency】、中位数【Median】、中位数检验【Median Test】、$t$ 检验【$t$-Test】。

## 参考文献


Goodman, A., Johnson, P., & Webb, S. (1997). *Inequality in the UK*. Oxford, UK: Oxford University Press.

Stuart, A., & Ord, J. K. (1987). *Kendall's advanced theory of statistics: Vol. 1. Distribution theory* (5th ed.). London: Griffin.

Yule, G. U., & Kendall, M. G. (1950). *An introduction to the theory of statistics* (14th ed.). London: Griffin.

# B

## 条形图(Bar Graph)

条形图是最为简单的演示离散值变量的相对频数或比例的方法之一。一般来讲,图中条形的长方形表示不同的值(如年龄的范围、教育程度、政治面貌和第一语言等)。条形自 $x$ 轴升起,其高度表示频数,频数的值可从 $y$ 轴上得知。我们也可以将条形设在 $y$ 轴上,而从 $x$ 轴上得知频数的读数。我们将这样设置的条形图叫作水平条形图。条形图以这种简单的形式所展示的信息,也可以用一个二维的饼图【Pie Chart】来展示。图 1 是一个条形图的实例,它给出了一组 100 个家庭的孩子数目的信息。我们从这张图中不难看出:(1)孩子的众数;(2)家庭规模较大的家庭频数比较低。

图 2 是一个水平条形图的例子。它显示了组成英国总人口的不同族群的比例。该图数据来自 1991 年的人口普查。

条形图也可以通过类别的归并或堆叠显示多个变量的信息。图 3 和图 4 在直观地显示了图 2 中族群信息的同时,引入了人口的年龄结构。图 3 是一个对类别进行了归并的聚类条形图实例,在该图中,每个列的百分比之和都为 100%。该图虽然能使我们得知不同族群有不同的人口结构,但却未能告诉我们不同族群人口的绝对数。而图 4 则与之不同,它将族群归并成两类:白人和所有的少数民族;分别将两类人口的表示不同年龄的条形排在一起,并标出实际的频数。

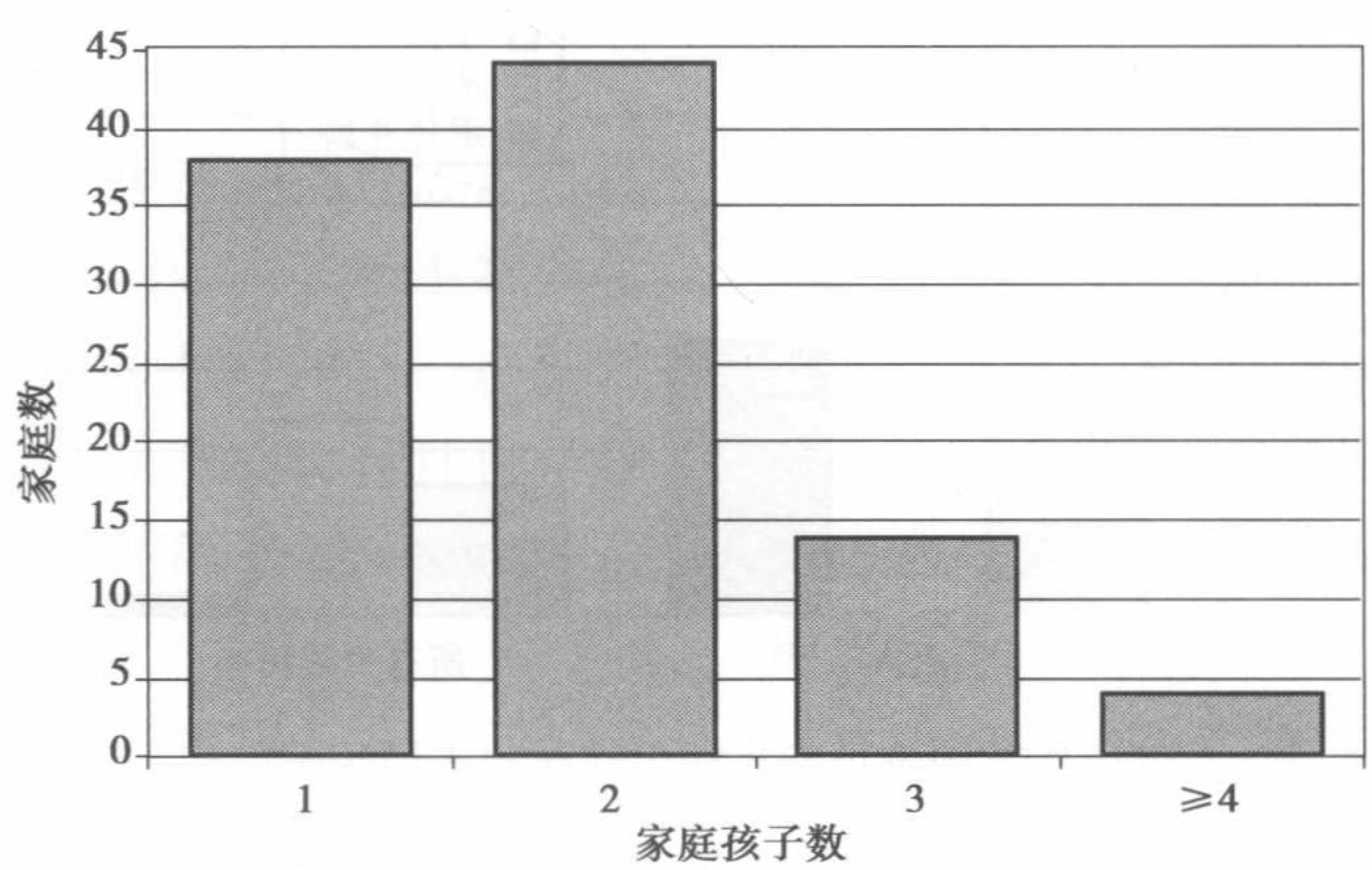



图 1　100 个样本家庭中,每个家庭的孩子数

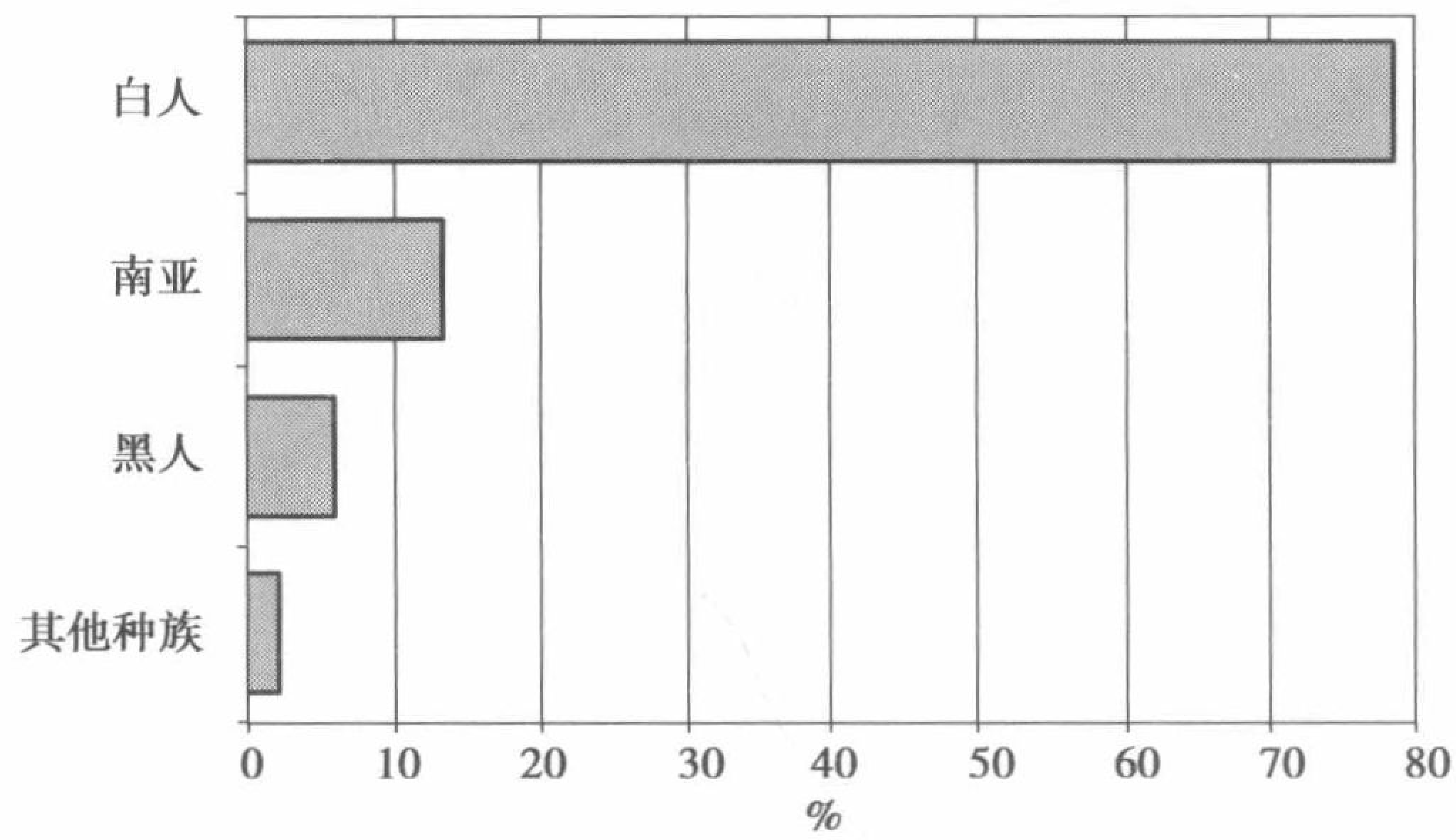



图 2　1991 年英国城市族群人口条形图

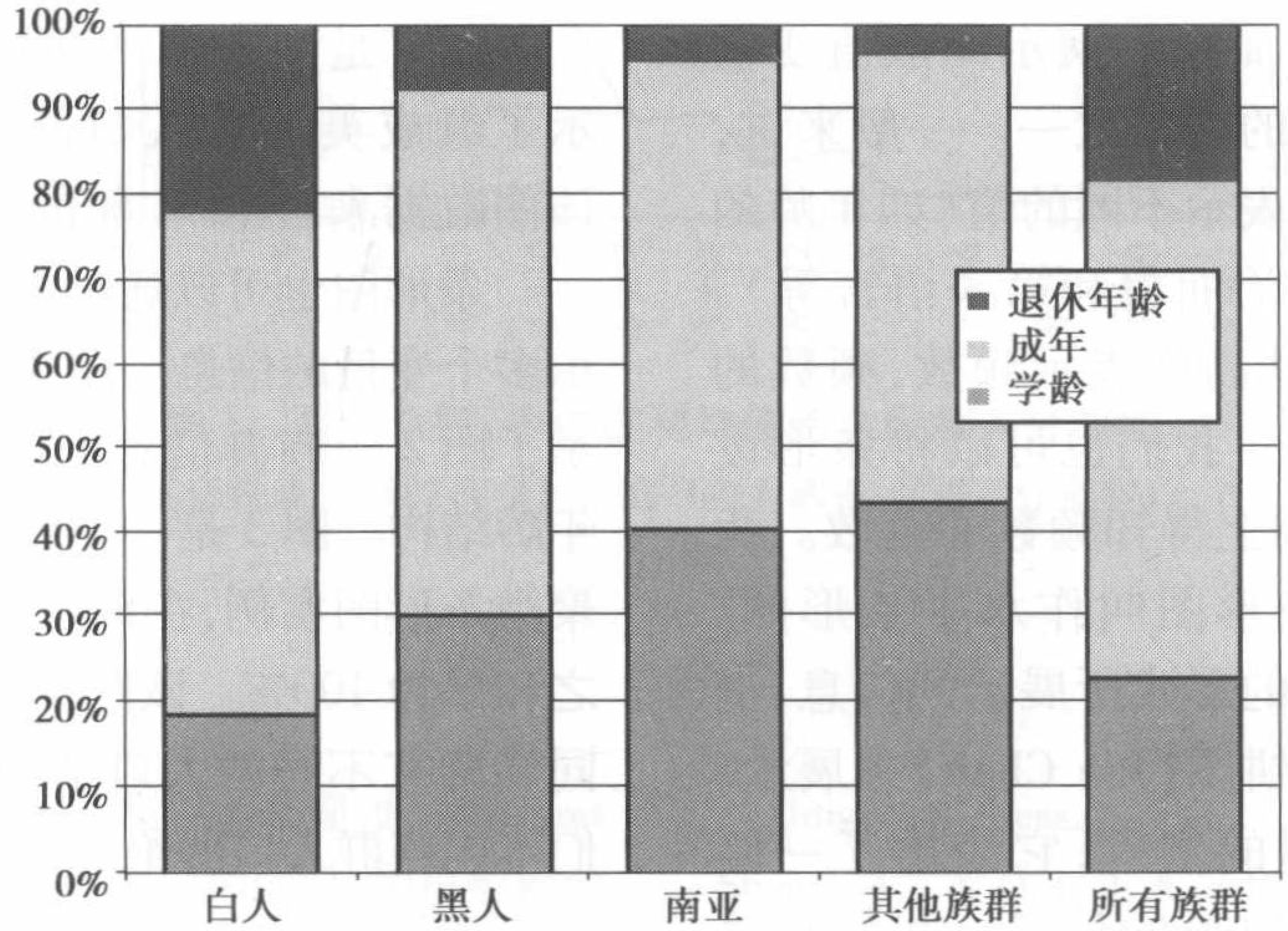



图 3　1991 年英国城市不同族群人口结构堆叠条形图

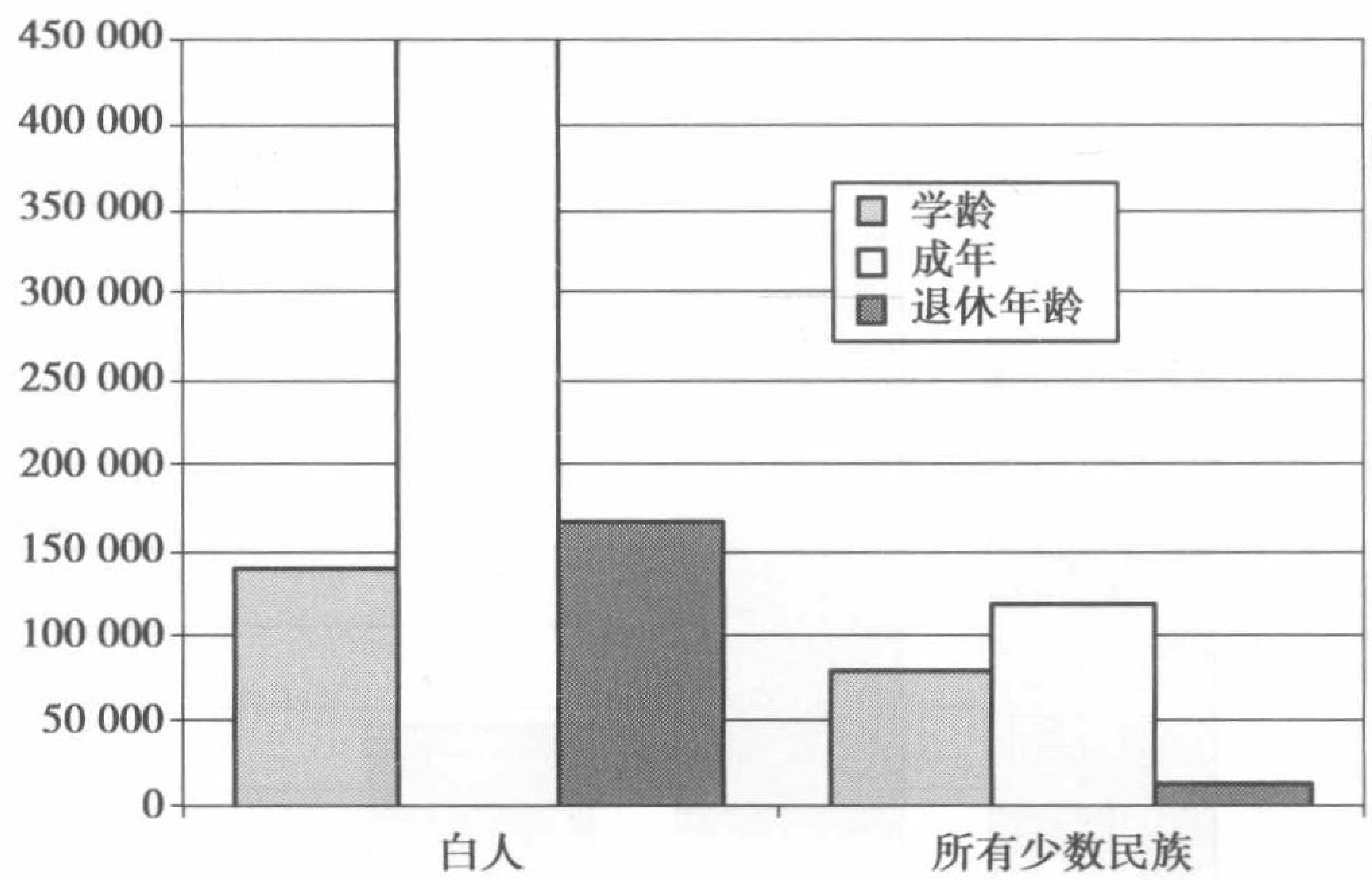



图 4　1991 年英国城市白人和少数民族人口年龄结构比较

我们之所以用条形图来概括信息，不仅是因为它明白易懂，更是因为如约翰 · 图基所言，它能引导我们更为深入地了解数据。查尔斯 · 普莱费尔在 18 世纪末，首先在统计和社会数据的分析中开发和应用了条形图，而在此之前的源流至今仍无从得知，这

确实令人惊讶。然而在今天,由于计算机软件的发展,创建更为复杂的三维堆叠和类别归并图都已不是什么难事。同样我们也可以轻而易举地对图形的外观、网格线及轴进行需要的调整。在社会科学中用条形图迅速而通达地描述数据和它们的特点不仅已经十分普遍,而且是为人们所热切期待的。即便如此,今天在绘制条形图时,究竟应该选择什么样的条形图(如条形图的形状、应该包含什么样的数据,以及什么应该形成轴、频数或比例等)仍然颇费思索。而作出什么样的选择可能最终取决于我们究竟怎样理解数据。

——Lucinda Platt

(沈崇麟译校)

### 参考文献

Tufte, E. R. (1983). *The visual display of quantitative information*. Cheshire, CT: Graphics Press.

Tukey, J. W. (1977). *Exploratory data analysis*. Reading, MA: Addison-Wesley.

Wallgren, A. et al. (1996). *Graphing statistics and data: Creating better data*. Thousand Oaks, CA: Sage.

## 基线(Baseline)

基线是用来比较研究发现的参照点。例如,在实施一个毒品预防的项目之前,研究者可以在目标总体中设立一个目前毒品使用的平均数作为评估项目效应的基线。在回归【Regression】分析中,当自变量【Independent Variable】$X=0$ 时,因变量【Dependent Variable】$Y$ 的期望值便是基线的又一个例子。因为这个值可以与 $X=1$,或其他非零值时 $Y$ 的期望值相比较。有时整个模型的结果也可作为与其他模型比较的基线。

——Michael S. Lewis-Beck

(沈崇麟译校)

## 基础研究(Basic Research)

好奇心是人类心理活动的标志。自古希腊开始,人们便一直认为,了解事物发生的原因,是一个基本的人类冲动。这种对知识的激情不仅产生了知识,而且也是我们日常生活巨大改进的基础(Kornberg, 1997)。

基础研究为了解世界运行方式的好奇心所驱动。或早或迟,新知识将给人们带来有实际利益的结果,有时甚至是令人震惊的利益结果。但这些利益结果不仅是无法预知的,而且有时的确也是难以直接寻找的。历史记录无可辩驳地证明,好奇心是发明创造的原动力。涉足于基础研究的科学家们,除了无时无刻不在想着证明自己之外,从不考虑利益和发明创造这样的事情。他们无时无刻不在受着自己的好奇心的驱使。但那些建构科学基础结构或掌握研究方法体系的人,常常可能会因为人们对兴利除弊问题的重视而被过度关注,尽管基础科学课程是不可或缺的(Kornberg, 1997)。

基础研究的特征可以以几种方式来表明。有些人认为,基础研究具有创新性和冒险性。但本质上说,基础研究的确是一种朴实的研究,其目的无非是了解某些现象和过程的运作方式。当然,现象和过程不仅在数量上各不相同,而且它们所涉及的领域也千变万化。例如,人类认识物理性质的一个高点是,认识到地球和天体现象受着同一个基本过程控制。正因如此,虽然基础研究为好奇心所驱使,且所有基础研究的目的都在探寻某些事物的运行方式,但过程越基础,研究就越基础。圣杯总是在于发现终极力量。虽然我们在这条路上投入了大量的精力和财力,然而过程确实是无处不在的,尽管我们渴望能达到尽可能远的地方,但我们到达的地方并非真正的终极力量之所在。沿着这条路进行的不懈探索,终究能使我们对世界的运行方式有所了解,与此同时,我们也在为我们尚未触及的过程的探索打磨科学的工具。

依照牛顿(Newton,1686/1952)理解物理本质的理论,可观察的行为和社会现象可被视为几种基本力量同时操作的产物。换句话说,社会行为世界是一个多因子世界。这一观点位于理论和经验研究工作的核心(Parsons, 1968)。科学的挑战有两种:一种是理论上的;另一种则是经验上的。理论挑战是确认控制人们行为的力量,描述它们的运作规律,并推导出它们的影响。经验挑战是检验推导出的影响。理论和经验研究工作一起导致有关人的行为和社会现象可靠的知识的积累。

即使最为简要的基础研究的思考,如果没有对控制行为和社会现象的基本力量的性质最低限度的思考,也不能被认为是一个完整的思考。费米曾经每天花一小时来苦思物理世界的本质(Segrè,1970)。本着这种精神,候选的基本力量初始清单可以包括姚肖(Jasso,2001, pp. 350-352)讨论的四个方面:

①了解事物的原因;

②判别事物的优度;

③完美;

④自由。

——Guillermina Jasso

(沈崇麟译校)

## 延伸阅读

应当要求所有的社会科学家、所有的学生、所有的院长、所有的论文评阅人和所有的研究和开发机构中的研究人员阅读科恩伯格(Kornberg,1997)的一篇小论文。

## 参考文献


Jasso, G. (2001). Rule-finding about rule-making: Comparison processes and the making of norms. In M. Hechter & K.-D. Opp (Eds.), *Social norms* (pp. 348-393). New York: Russell Sage.

Kornberg, A. (1997). *Basic research, the lifeline of medicine.* Retrieved from http://www.nobel.se/medicine/articles/research/.

Newton, I. (1952). *Mathematical principles of natural philosophy.* Chicago: Britannica. (Originally published in 1686)

Parsons, T. (1968). Émile Durkheim. In D. L. Sills (Ed.), *International encyclopedia of the social sciences, Vol. 4.* New York: Macmillan.

Segrè, E. (1970). *Enrico Fermi: Physicist.* Chicago: University of Chicago Press.

## 贝叶斯因子(Bayes Factor)

已知数据 $y$ 和两个模型 $M_0$ 和 $M_1$,贝叶斯因子

$$B_{10}=\frac{p(y\mid M_1)}{p(y\mid M_0)}$$
$$=\left\{\frac{p(M_1\mid y)}{p(M_0\mid y)}\right\}\Bigg/\left\{\frac{p(M_1)}{p(M_0)}\right\}$$

是由数据提供的 $M_1$ 对 $M_0$ 的证据的概括。同时它也是后与前可能性比(odds)之比(故称为"贝叶斯因子")。2 倍的 $B_{10}$ 的对数与模型比较的偏差和似然比检验统计量是等尺度的。杰弗里(Jeffreys,1961)建议在解释贝叶斯因子时用下面的尺度:

| $B_{10}$ | $2\log B_{10}$ | $M_1$ 的证据 |
|---|---|---|
| <1 | <0 | 负(支持 $M_0$) |
| 1~3 | 0~2 | 几乎不值一提 |
| 3~12 | 2~5 | 正 |
| 12~150 | 5~10 | 强 |
| >150 | >10 | 很强 |

注意,$B$ 是边缘似然比,它等于 $p(y|M_k)=\int p(y|\theta_k,M_k)p(\theta_k)\mathrm{d}\theta_k$,式中的 $\theta_k$ 是模型 $M_k$ 的参数向量,$p(\theta_k)$ 是 $\theta_k$ 的先验密度,而 $p(y|\theta_k,M_k)$ 则是在模型 $M_k,k\in\{0,1\}$ 下的 $y$ 的似然。拉夫特里(Raftery,1996)提供了一个简要的计算贝叶斯因子的方法。而古德(Good,1988)则概括了贝叶斯因子的历史。

——Simon Jackman

(沈崇麟译校)

*也可参见贝叶斯定理【Bayes' Theorem】。

### 参考文献

Good, I. J. (1988). The interface between statistics and philosophy of science. *Statistical Science*, *3*, 386-397.

Jeffreys, H. (1961). *Theory of probability* (3rd ed.). Oxford, UK: Clarendon.

Raftery, A. E. (1996). Hypothesis testing and model selection. In W. R. Gilks, S. Richardson, & D. J. Spiegelhalter(Eds.), *Markov chain Monte Carlo in practice* (pp. 163-187). London: Chapman and Hall.

## 贝叶斯定理,贝叶斯法则(Bayes' Theorem, Bayes' Rule)

贝叶斯统计和决策理论的核心是贝叶斯定理,它也常被称为贝叶斯法则。它的最简单的形式可表述为:如果 $H$ 和 $E$ 分别表示假设和证据,那么贝叶斯定理便是:

$$\Pr(H\mid E)=\frac{\Pr(E\cap H)}{\Pr(E)}$$
$$=\frac{\Pr(E\mid H)\Pr(H)}{\Pr(E)}$$

如果 $\Pr(E)>0$,那么 $\Pr(H|E)$ 就是在得到 $E$ 之后相信 $H$ 的概率【Probability】,而 $\Pr(H)$ 则是一个在考虑 $E$ 之前的先验概率。定理的左方,$\Pr(H|E)$ 通常叫作 $H$ 的后验概率。这样,定理就提供了一种解决一般的推论或归纳问题的方法(Hacking, 2001),同时,给了我们一种从数据 $E$ 了解假设 $H$ 的途径。

贝叶斯定理本身不存在任何争议:它无非是一个由基本原理推出的一个会计恒等式(accounting identity),将联合、条件和边缘概率连接起来,也即 $\Pr(A \mid B)\Pr(B) = \Pr(A \cap B)$,式中的 $\Pr(B) \neq 0$。因此贝叶斯定理有时也叫作逆概率法则,因为它揭示了怎样可以把一个给定 $A$ 的条件概率"倒转",产生一个给出了 $B$ 的条件概率 $A$(Leamer, 1978, p. 39)。此外,定理还给我们提供了依据优势比【Odds Ratio】来思考两个假设 $H_i$ 和 $H_j$ 的途径,它也遵循贝叶斯定理,即

$$\frac{\Pr(H_i \mid E)}{\Pr(H_j \mid E)} = \frac{\Pr(E \mid H_i)\Pr(H_i)}{\Pr(E \mid H_j)\Pr(H_j)}$$

令后验的 $H_i$ 比 $H_j$(已知 $E$)的优势比率等于 $H_i$ 和 $H_j$ 下的 $E$ 的似然比乘以先验可能性。在连续情形中(如了解已知数据 $y$ 的实值参数 $\theta$)贝叶斯定理常被表达为:

$$\pi(\theta \mid y) \propto f(y;\theta)\pi(\theta)$$

式中的 $f(y;\theta)$ 是似然函数,比例常数 $[f(y;\theta)\pi(\theta)d\theta]^{-1}$ 并不取决于 $\theta$,并确保后验密度的积分为 1。这一版本的贝叶斯定理对贝叶斯统计是很重要的,它揭示了似然函数(已知 $\theta$ 数据的概率)怎么能在已知数据 $y$ 时,转化为关于 $\theta$ 的概率陈述。读者也要注意,贝叶斯法将参数 $\theta$ 也看成一个随机变量和数据的一些条件,而频率论者采用的方法则是将 $\theta$ 看成我们从中随机地抽取样本数据 $y$ 的总体的一个固定的(但是未知的)性质。

雷韦朗·托马斯·贝叶斯死于 1761 年。他的研究成果,现在被称为贝叶斯定理出现在一篇被认为由他撰写的文章中。在贝叶斯去世之后,该文于 1763 年由理查德·普莱斯递呈皇家学会(Bayes, 1763),以后多次再版(如 Bayes, 1958)。贝叶斯本人阐明,结果只针对一致的先验(uniform prior)。根据施蒂格勒(Stigler, 1986b)所述,拉普拉斯于 1774 年(显然在不知道贝叶斯的著作的情况下)以更为一般的形式(针对离散事件)阐述了这一定理。有关贝叶斯定理的更为详细的历史可参见伯纳多和史密斯(Bernardo & Smith ,1994, Chap. 1)、林德利(Lindley, 2001)和施蒂格勒(Stigler, 1986a, Chap. 3)等人的有关著作。

——Simon Jackman

(沈崇麟译校)

## 参考文献


Bayes, T. (1763). An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society*, *53*, 370-418.

Bayes, T. (1958). An essay towards solving a problem in the doctrine of chances. *Biometrika*, *45*, 293-315.

Bernardo, J., & Smith, A. F. M. (1994). *Bayesian theory.* Chichester, UK: Wiley.

Hacking, I. (2001). *An introduction to probability and inductive logic*, Cambridge, UK: Cambridge University Press.

Leamer, E. (1978). *Specification searches: Ad hoc inference with nonexperimental data.* New York: Wiley.

Lindley, D. V. (2001). Thomas Bayes. In C. C. Heyde & E. Seneta (Eds.), *Statisticians of the centuries* (pp. 68-71). New York: Springer-Verlag.

Stigler, S. (1986a). *The history of statistics: The measurement of uncertainty before* 1900. Cambridge, MA: Belknap.

Stigler, S. (1986b). Laplace's 1774 memoir on inverse probability. *Statistical Science*, *1*, 359-378.

## 贝叶斯推论(Bayesian Inference)

贝叶斯统计法是一种用雷韦朗·托马斯·贝叶斯在1763年的著名论文中提出的概率【Probability】定理,是一种以更新的信息为依据的统计法。贝叶斯是一位业余的数学家,他的论文是在他去世后两年,被他的朋友理查德·普莱斯发现并付梓的。和他的名字相连的一个统计学分支之所以能长盛不衰,完全是因为他所作的概率陈述的重要性。现在人们把它叫作贝叶斯定律或贝叶斯定理【Bayes' Theorem】,它与条件概率密切相关。

肇端于两个感兴趣的非独立的事件,$X$和$Y$。由概率的基本定理可知,若$Y$已经发生,那么$X$的条件概率可由下式

$$p(X\mid Y)=\frac{p(X,Y)}{p(Y)}$$

求得。式中的$p(X, Y)$是“$X$和$Y$二者都发生的概率”,而$p(Y)$是$Y$发生的无条件概率。因而条件表达式给出了$Y$的某个事件发生之后,$X$发生的概率。我们也可以定义与此不同的,即$X$首先发生的条件概率,

$$p(Y\mid X)=\frac{p(Y,X)}{p(X)}$$

因为$X$和$Y$发生的联合概率与$Y$和$X$同时发生的联合概率$[p(X,Y)=p(Y,X)]$相同,所以我们可以将表达式重新整理简化为:

$$p(X,Y)=p(X\mid Y)p(Y)$$
$$p(Y,X)=p(Y\mid X)p(X)$$
$$p(X\mid Y)p(Y)=p(Y\mid X)p(X)$$
$$\text{所以}\ p(X\mid Y)=\frac{p(X)}{p(Y)}p(Y\mid X)\qquad(1)$$

(更为详尽的阐述,请参见Bernardo & Smith, 1994, Chap. 3)上面最后一行便是著名的贝叶斯定理,而这一定理便是一种真正可令条件概率“换向”的方法。问题很清楚,我们只需将上面最后一个等式中的无条件概率挪到等式的左面,便可轻而易举地得到$p(Y|X)$。贝叶斯定理是很有用的,因为我们可能经常知道$p(X|Y)$,并且很想知道$p(Y|X)$,或者恰好相反。而$p(X|Y)$永远也不会等于$p(Y|X)$(即

$$\frac{p(X)}{p(Y)}=1$$

的概率为零)。这一事实常被称为逆概率问题。

而这一简单的结果正是贝叶斯统计的一般范式的基础。

### 贝叶斯推论述要

贝叶斯推论依据的有关数据和参数的假定与经典统计学完全不同(Box & Tiao, 1973, Chap. 1)。在贝叶斯那里,所有的数量都被分成了两个部分:可观察的和不可观察的。可观察的数量一般都是数据和任何有关的常量。不可观察的数量包括要估计的感兴趣的参数【Parameters】、缺失数据【Missing Data】、只需解释一下的不太感兴趣的参数。所有可观察的数量都是固定的和有条件的。所有不可观察的数量都假定是拥有某种分布的数量,因此它们都可被作为随机变量处理。因此,在贝叶斯那里,不仅参数不再被作为全体总体中固定和不变动的数量来处理,而且所有的陈述也都使用概率的语言作出。

推论的过程开始于为未知参数设定先验分布【Prior Distribution】。这些设定的依据既可能是以前研究的信息量非常大的描述,也可能是对该领域不甚了了但却是经过慎重考虑的估计和猜测。先验分布并非处

理未知的数量强加的一种麻烦，相反，它是一个将定性的、记叙的和直觉的知识系统地引进统计模型的机会。下一个步骤是通过设定数据的参数形式和插入观察的数量，以常规方式拟定一个似然函数。最后一步是生成一个后验分布【Posterior Distribution】，方法是将先验分布和似然函数相乘。因此，这一似然函数要用数据来有条件地更新先验的知识。

这一过程可以简单地用助记符概括为：

后验概率 ∝ 先验概率 × 似然函数 (2)

这恰好就是一种贝叶斯定理的形式。在这种表达式中，右侧的分母，由于比例性的使用而被忽略。符号“∝”表示“成比例”，该符号意味着式中已无常数，这使后验和或积分将像概率质量函数和概率密度函数【Probability Density Function】要求的那样为1。我们总是可以在以后通过重新正态化得到一个适当的形式，再加上比例性的使用会使问题变得更为直观，从而大大减轻计算的负担。而上面的公式显示，后验分布是反映以往研究成果的先验分布和我们掌握的数据导出的似然函数之间的折中。贝叶斯推论可概括为以下三个一般化的步骤：

①设定一个概率模型，如果可能，其中可包括一些未知参数的先验知识；

②以观察数据上的这一概率模型为条件，更新有关未知参数的知识；

③评估模型与数据的拟合和结论对假定的敏感性。

在这一过程的任何地方都有基于某些没有效应的虚无假设【Null Hypothesis】为真的人为的决策。因此与有严重缺陷的虚无假设的显著性检验【Significance Test】不同，在这里，证据的递呈，只是通过对这种后验分布的概括来实现的。概括通常都用分位数或概率来陈述。譬如感兴趣的概率小于/大于某些感兴趣的常数，或这种参数占据某一区域的概率等。大家也要注意，如果后验分布先作为新的先验分布来处理，新的数据被观察到的话，那么它也同样是可以被更新的。因此有关感兴趣的参数的知识是随时间的推移而在不断被更新和积累的(Robert, 2001)。

## 似然函数

假设收集到的数据被作为一种固定的量来处理，且我们知道描述数据产生过程的恰当的概率质量函数或概率密度函数。标准的似然和贝叶斯法在某些方面是相似的：它们都始于这两个假定，然后都在一个参数模型中来求未知的参数值。最大似然估计【Maximum Likelihood Estimation】，用贝叶斯定理起始，以无限的似然概念替代了有限的概率定义：

$$p(\theta \mid \mathbf{X}) = \frac{p(\theta)}{p(\mathbf{X})} p(\mathbf{X} \mid \theta) \qquad (3)$$

式中的 $\theta$ 是感兴趣的未知的参数，而 $\mathbf{X}$ 则是收集到的数据。关键在于将

$$\frac{p(\theta)}{p(\mathbf{X})}$$

处理为独立于 $p(\mathbf{X}|\theta)$ 的数据的未知的函数。这将允许我们使用 $L(\theta|\mathbf{X}) \propto p(\mathbf{X}|\theta)$。因为数据是固定的，所以似然函数的不同的值只能通过插入未知参数 $\theta$ 的不同的值得到。

似然函数 $L(\theta|\mathbf{X})$ 虽然与我们希望的类似，但是其逆函数 $p(\mathbf{X}|\theta)$ 却难以得到，而它正是我们据之进行 $\theta$ 备择值检验，找到最可能的 $\hat{\theta}$ 所需要的。然而，因为似然函数已经不再受0和1的限制，所以现在它之所以重要只是相对于其他基于不同的 $\theta$ 值的似然函数而言。注意，先验 $p(\theta)$ 在这里基本上是被忽略的，而非完全解决了。这等价于在贝叶斯环境中设定一个统一的先验，一个观

察值，这使得有人认为经典的推论无非是贝叶斯主义的一个特例："人人都是贝叶斯，只是有些人知道它而已。"

一般，我们的兴趣在于得到 $\theta$ 的最大似然估计值。无约束的和未知的参数 $\theta$ 的值，提供了似然函数 $L(\theta|\mathbf{X})$ 的最大值。$\theta$ 的这一值记作 $\hat{\theta}$。如果 $H_0$ 是通过一个相对于 $\theta$ 的样本空间中其他那些可能值的参数形式来表达的，那么这个值就是最有可能产生这个数据的。

## 贝叶斯理论

贝叶斯法讨论的是逆概率问题。它先于观察数据 $\mathbf{X}$ 作出有关参数 $\theta$ 的无条件分布的分布假定：$p(\theta)$。先验和似然通过贝叶斯定理联系起来：

$$\pi(\theta|\mathbf{X}) = \frac{p(\theta)L(\theta|\mathbf{X})}{\int_{\Theta} p(\theta)L(\theta|\mathbf{X})\mathrm{d}\theta} \quad (4)$$

式中的 $\int_{\Theta} p(\theta)L(\theta|\mathbf{X})\mathrm{d}\theta = p(\mathbf{X})$。这里使用的记号 $\pi(\ )$ 主要是为了区别 $\theta$ 的后验分布和先验分布。分母上的那一项一般对推论不太重要，且可以在以后通过积分重新恢复。这一项一般叫作正态化常数、正态化因子或先验预测分布，尽管它实际上正好是数据的边缘分布，且确保 $\pi(\theta|\mathbf{X})$ 的积分为 1。

我们通过去掉分母和使用比例记号，得到(4)的一种更为简洁且有用的形式，因为 $p(\mathbf{X})$ 并不取决于 $\theta$，因而它并没有提供任何有关 $\theta$ 可能值的推论信息：

$$\pi(\theta|\mathbf{X}) \propto p(\theta)L(\theta|\mathbf{X}) \quad (5)$$

这意味着，感兴趣的参数未正态化的后验（抽样）分布，与先验分布和似然函数的乘积成比例。

最大似然估计等于一个有恰当的一致先验概率的贝叶斯后验模式，且它们都渐近于任何给定的先验（值）。二者在极限情况下都呈正态分布【Normally Distributed】。在许多情况下，先验的选择不是太重要的，因为随着样本量的增大，似然会逐渐主导先验。尽管贝叶斯定理将一个先验的分布给未知参数的做法被认为是主观的（虽然所有统计模型实际上都是主观的），但仍会有对一些特定的先验方式的强有力的论证：有限和模糊的知识常常会证明一个扩散甚至概率先验；某种概率模型会在逻辑上导致先验的某种特定形式（共轭性）和先验能使研究者将当前研究之外收集的其他信息包括进来。

## 总结贝叶斯结果

使用贝叶斯法的研究者一般不关心只是得到感兴趣的参数 $\theta$ 特定的点估计，将它作为一种提供概率分布中经验证据的方法。相反，他们的注意力都集中在描述 $\theta$ 的后验分布的形状和特点上。如描述直接概率区间的一般形式（可靠集合和最高的后验密集区域），后验分位数和诸如 $p(\theta_i<0)$ 这样感兴趣的概率。

在贝叶斯设置中，我们也可以进行假设检验。假如 $\Theta_1$ 和 $\Theta_2$ 表示两个竞争的、陈述某个未知参数 $\theta$ 的假设，并形成了样本空间 $\Theta=\Theta_1\cup\Theta_2$，$\Theta_1\cap\Theta_2=\phi$ 的一部分。先验概率分配给了两个结果中的每一个：$\pi_1=p(\theta\in\Theta_1)$ 和 $\pi_2=p(\theta\in\Theta_2)$。这将从两个先验和似然函数导致两个竞争的后验分布：$p_1=p(\theta\in\Theta_1|\mathbf{X})$ 和 $p_2=p(\theta\in\Theta_2|\mathbf{X})$。一般都将先验的可能性比 $\pi_1/\pi_2$ 和后验的可能性比 $p_1/p_2$ 作为 $H_1$ 和 $H_2$ 的证据。而一个更为有用的数量是 $(\pi_1/\pi_2)/(p_1/p_2)$，叫作贝叶斯因子【Bayes Factor】。贝叶斯因子常被解释为优比，即在给定观测数据的条件下，命题 $H_1$ 比 $H_2$ 更可信的程度。正因如此，它自然会导向假设检验的贝叶斯模拟。

## 例子

假设 $x_1$, $x_2$, $\cdots$, $x_n$ 是独立的随机变量，它们都是由有参数 $\theta$ 的伯努利质量函数产生的，所以它们的和

$$y = \sum_{i=1}^{n} x_i$$

服从有已知 $n$ 和未知 $\theta$ 的二项分布$(n,\theta)$。将一个 beta$(A,B)$分配给 $\theta$ 的先验分布，

$$p(\theta \mid A,B) = \frac{\Gamma(A+B)}{\Gamma(A)\Gamma(B)}\theta^{A-1}(1-\theta)^{B-1}$$

$$0 < \theta; 0 < A \text{ 或 } B$$

$A$ 和 $B$ 的值由研究者设定。

$\theta$ 的后验分布用贝叶斯法则求得，

$$\pi(\theta \mid y) = \frac{p(\theta \mid A,B)p(y \mid \theta)}{p(y)} \quad (6)$$

式中的 $p(y|\theta)$是 $y$ 的概率质量函数。现在，右侧的分子很容易计算，呈假定的参数形式：

$$\begin{aligned} p(y,\theta) &= p(y \mid \theta)p(\theta) \\ &= \left[\binom{n}{y}\theta^{y}(1-\theta)^{n-y}\right] \times \left[\frac{\Gamma(A+B)}{\Gamma(A)\Gamma(B)}\theta^{A-1}(1-\theta)^{B-1}\right] \\ &= \frac{\Gamma(n+1)\Gamma(A+B)}{\Gamma(y+1)\Gamma(n-y+1)\Gamma(A)\Gamma(B)} \times \theta^{y+A-1}(1-\theta)^{n-y+B-1} \end{aligned} \quad (7)$$

而分母则可通过求这一联合分布的 $\theta$ 的积分求得：

$$\begin{aligned} p(y) &= \int_0^1 \frac{\Gamma(n+1)\Gamma(A+B)}{\Gamma(y+1)\Gamma(n-y+1)\Gamma(A)\Gamma(B)} \times \theta^{y+A-1}(1-\theta)^{n-y+B-1}\mathrm{d}p \\ &= \frac{\Gamma(n+1)\Gamma(A+B)}{\Gamma(y+1)\Gamma(n-y+1)\Gamma(A)\Gamma(B)} \times \frac{\Gamma(y+A)\Gamma(n-y+B)}{\Gamma(n+A+B)} \end{aligned} \quad (8)$$

更为详尽的技术细节请参见吉尔(Gill, 2002, p. 67)进行的一系列数学运算，算式(6)将给我们提供

$$\begin{aligned} \pi(\theta \mid y) &= \frac{p(\theta)p(y \mid \theta)}{p(y)} \\ &= \frac{\Gamma(n+A+B)}{\Gamma(y+A)\Gamma(n-y+B)} \times \theta^{(y+A)-1}(1-\theta)^{(n-y+B)-1} \end{aligned} \quad (9)$$

这是一个基于数据 $y$ 的更新先验分布的，有参数 $A'=y+A$ 和 $B'=n-y+B$ 的 $\theta$ 的新的 beta 分布。因为先验和后验有着相同的分布族形式，所以这是一个有趣的特例，叫作共轭模型(beta 分布是对二项概率质量函数共轭的)。共轭模型是简约的，但是在一般情况下这会使实际的贝叶斯的统计设定变得很复杂，有时甚至是不可能的。

## 马尔科夫链蒙特卡洛法

在 20 世纪，对于那些开发贝叶斯模型的统计学家来讲，一个始终缠绕着他们的问题是，经常有可能从先验和似然相乘得到足够多的复杂的后验，虽然数学形式存在，但是却无法计算可用于分析的感兴趣的数量，如均值和分位数。这一问题在相当长的一段时间内，使贝叶斯方法被边缘化于主流统计学，这样的情况直至盖尔范德和史密斯(Gelfand & Smith, 1990)的综述性论文的发表才有所改观。该论文阐述了在统计物理学中，类似的问题是如何用马尔科夫链模拟技术来解决的。基本上

讲,这种技术是一种迭代技术,它用感兴趣的后验(逐渐地)产生值(Gill, 2002)。因此,不同的后验形式可以用大量的模拟值在经验上描述,因此,通过模拟,可进行不同的积分计算。这一发展的结果,导致解决很多未曾解决的贝叶斯统计问题的论文的涌现,而其对贝叶斯统计的影响无疑可被视为一场革命。

马尔科夫链由那些连续的,在概率上只取决于它们的直接前导的数量构成。一般来讲,我们有可能设置一条估计多维度概率结构,希望得后验分布的链,其方法是使马尔科夫链在一个适当的样本空间中开始,让它运行至适应于一个目标分布为止。然后,在它只限于为这一特定的分布运行一段时间之后,它可能从模拟值中收集到概括性统计值,如均值、方差和分位数。两种使用最为普遍的方法是 Metropolis-Hastings 算法和 Gibbs 抽样【Gibbs Sampling】。已经证明,它们都拥有从目标分布中导出经验样本的理论属性。这些方法不仅合理,且简单易行,因而在社会科学中使用得越来越普遍。

——Jeff Gill

(沈崇麟译校)

## 参考文献


Bernardo, J. M., & Smith, A. F. M. (1994). *Bayesian theory*. New York: Wiley.

Box, G. E. P., & Tiao, G. C. (1973). *Bayesian inference in statistical analysis*. New York: Wiley.

Gelfand, A. E., & Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, *85*, 389-409.

Gill, J. (2002). *Bayesian methods: A social and behavioral sciences approach*. New York: Chapman and Hall.

Robert, C. P. (2001). *The Bayesian choice: A decision theoretic motivation* (2nd ed.). New York: Springer-Verlag.


## 贝叶斯模拟(Bayesian Simulation)

参见马尔科夫链蒙特卡洛法【Markov Chain Monte Carlo Methods】。

## 行为编码(Behavior Coding)

行为编码是一种检验调查访谈员和被访者之间互动的系统、灵活、高效的办法。编码员倾听现场访谈或访谈的录音,给各种行为进行编码【Codes】,诸如访谈员宣读问题的准确性和被访者回答问题的困难程度。这种编码结果可以用于监测访谈员的表现。尽管这些目标都需要有比较复杂的程序,但大体的过程是相同的。研究者必须确认要对什么行为进行编码,开发一个合适的编码方案,挑选和培训编码员,对信度进行评估,监督编码任务,最后还要对结果进行分析。

自 1960 年代末开始,调查方法专家曾以访谈录音的行为编码为根据,对访谈做了各种各样的研究(Cannell, Fowler, & Marquis, 1968)。这些研究对调查访谈的各种要件都提供了有价值的信息,包括开放式问题【Open-Ended Questions】和封闭式问题

【Closed-Ended Questions】的行为差异、不同特征受访者的行为差异,以及有经验的访谈员和无经验的访谈员的行为差异。这些研究中最有价值的发现便是调查并非是按研究者预期的那样进行的。问题没有按照规定的方式来问,访谈中也没有遵循调查说明规定的其他技术规范。最初我们认为这都是访谈员之过,将访谈中的各种问题归咎于训练不足、缺乏监督,或仅仅是访谈员偷懒。但是,经过仔细分析之后,我们发现大部分困难都应归结于对访谈员进行培训的技术或是调查的问题中存在的不足。

不必奇怪,调查研究中行为编码的最早的实际例行应用就是监督访谈员的表现。当今行为编码已经是调查问题的设计阶段所采用的技术,而且是许多机构在预调查期间评估问题使用的一种常规技术(Oksenberg, Cannell, & Kalton, 1991)。大多数为评估调查的问题而设计的编码方案既要考察访谈员最初如何宣读问题,也要对所有后来的受访者的行为进行考察。访谈员宣读的编码评估的是(访谈员)按照问题的措辞照本宣科的准确程度。而受访者行为编码则试图在较低限度上捕捉与理解困难相关的行为(例如,要求解释清楚)、与问题措辞或长度相关的潜在的问题(例如,各种中断)、答案的恰当性(例如,不当的答案、合格的答案、不充分的答案),以及不能回答的情况(例如,答案为“不知道”或者拒绝回答)。“差”的行为出现频率高的话,表明问卷措辞或是答案选项中可能存在问题。

——Nancy Mathiowetz

(杨可译　沈崇麟校)

*也可参见访员效应【Interviewer Effects】、访员培训【Interviewer Training】、前测【Pretesting】。

## 参考文献

Cannell, C., Fowler, J., & Marquis, K. (1968). *The influence of interviewer and respondent psychological and behavioral variables on the reporting of household interviews* (Vital and Health Statistics Series 2, No. 6). Washington, DC: National Center for Health Statistics.

Oksenberg, L., Cannell, C., & Kalton, G. (1991). New strategies for pretesting survey questions. *Journal of official Statistics*, *7*(3), 349-365.

# 行为科学(Behavioral Sciences)

行为科学关注的是对人类行为的解释。心理学是最主要的例子,但其他的社会科学也有行为方面的分支。例如,政治学中对投票行为的研究以及经济学里对消费者行为的研究都很发达。

——Michael S. Lewis-Beck

(杨可译　沈崇麟校)

# 行为主义(Behaviorism)

行为主义是约翰·华生(Watson, 1959)发明的一个术语,它既是一种方法,又是一种形而上学。以第一人称报告的自我意识是不可信的,因为它们是无法独立查证的。有鉴于此,方法论行为学家试图以他人对某一对象的行为的对照观察来代替个人

对其心态的内省。行为学家不会把一个愤怒的小孩对其情绪的描述当作确定的,而是会去关注她是否试图损坏那些据说惹恼了她的东西。依此理由,我们认为,仅由一人以独有的方式观察到的事物是不真实的。形而上的行为论者坚持认为,情绪和其他所谓的心灵状态都不是私密所有,仅能由据有它们的人或动物所知晓,而是众人所能观察到的,以确定方式表现于行为的性情。因此,形而上的行为学家主张,因为愤怒一般总是存在于一种毁坏东西的性情中,所以我们可以观察一个小孩的行为表现来观察他/她的情绪。正是由于这样的定义,对行为主义持批评态度的人用一个笑话给行为主义描绘了一幅惟妙惟肖的漫画,用它来述说两个行为主义者在街上相遇时的情形——一个行为主义者对另一个说:"你感觉不错,我感觉怎么样?"

内省这种第一人称的方法,即我们所谓的(人类的)意识现象,在18、19世纪主导着心理学。20世纪的大多数时间里,精心控制条件下行为观察(包括动物和人类)的第三人称方法主导着所谓的行为科学。不过,到了20世纪末,许多认知科学家逐渐认识到,行为学家以前对内省和意识的责难过于狭隘。在这些新的观点看来,感觉、情绪、想法以及其他的心态确实都是大脑的状态。这些状态一般来说会导致各种特异的行为,但作为原因,它们与行为的关联却与他们的效果相关。因此,有人主张,这些心态的本质不能单凭观察它导致的行为来发现,正如我们不能通过观察磁铁吸引铁粉就发现其物理结构一样。根据人类大脑有监控其自身状态的特殊能力这一理论,有些科学家和哲学家又开始研究"意识"这个19世纪的哲学家最喜欢的主题,把内省和大脑生理学研究结合起来。行为主义者对这种老项目的更新作出了回应。他们的回应并非否定将研究与脑生理学关联,而是坚持认为所谓的意识现象是虚幻的,而非过硬的数据(hard data),正如哈沃德·拉克林试图证明的,行为才是理解心智秘密的关键所在。

——Max Hocutt

(杨可译　沈崇麟校)

## 参考文献


Rachlin, H. (2000). *The science of self-control.* Cambridge, MA: Harvard University Press.

Watson, J. B. (1959). *Behaviorism.* Chicago: University of Chicago Press.


# 钟形曲线(Bell-shaped Curve)

钟形曲线这一术语是专门用来描述正态分布的曲线形状的。这种曲线的 $x$ 轴显示分布不同的值,而 $y$ 轴则显示各个不同的值出现的频数。钟形曲线是一种对称分布,因此最高的频数都围绕在分布中点的周围,在频数离开中心时,分布的两侧都以相等速率向0逐渐拖尾。图1显示了一个诸如这样的钟形曲线。而由对称性和曲线的形状可见,所有三个集中趋势量度【Measures of Central Tendency】——均值、众数和中位数同时发生在曲线的最高点。

我们用钟形曲线的均值 $\mu$ 和标准差【Standard Deviation】$\sigma$ 来描绘它。每个有特定的 $\mu$ 和 $\sigma$ 的钟形曲线将代表一个唯一的分布。因为均值是趋向曲线的中央变大和围绕在均值周围的,所以任何来自钟形曲线

分布的单个观察值落入均值附近的概率,远远大于落入任何一个尾端的概率。其结果诚如我们由正态概率所知,在钟形曲线中,将有68%的值落入距均值一个标准差的范围内,而将有95%的值落入距均值大约两个标准差的范围内。几乎所有的值都将落入距均值三个标准差的范围内。其余的观察将形成两个尾部。这在图1中得到了直观的显示。在图1中我们可以看出,2.5%的个案落在了$\mu \pm 1.96 \times \sigma$之外,即钟形曲线的尾部。这一数字给出了近似的钟形抽样分布的概率,任何个案都有95%的概率落入这一范围内。因此我们可以用统计的外推法来预测总体参数,因为它落入这样的范围内有置信度95%(参见中心极限定理【Central Limit Theorem】、正态分布【Normal Distribution】)。

一个均值为0,标准差为1的特定形式的钟形曲线叫作标准正态分布。一个诸如其分布概率由等式$\mu \pm z^{*}\sigma$表示的分布可简化为$\sigma$(或$Z$得分)的倍数的值。因此,对一个标准的正态分布而言,95%的概率位于±1.96这一范围之内。

于是问题很清楚,钟形分布的对称性将导出某种我们可能据之对任何近似于这种形状的分布做统计推论的特别的统计性质。即使在分布并非钟形分布时,借助外推法,钟形分布也仍然可以构成统计推论的基础。此外,非钟形分布的分布可通过转换来创建近似的钟形分布曲线。例如,收入分布,其分布呈右向的倾斜,我们可取收入值的对数把它转换成一种近似于对称的分布。相反,一个左倾的分布,如考试成绩或期望寿命则可通过取值的平方或立方转换成近似的对称分布。

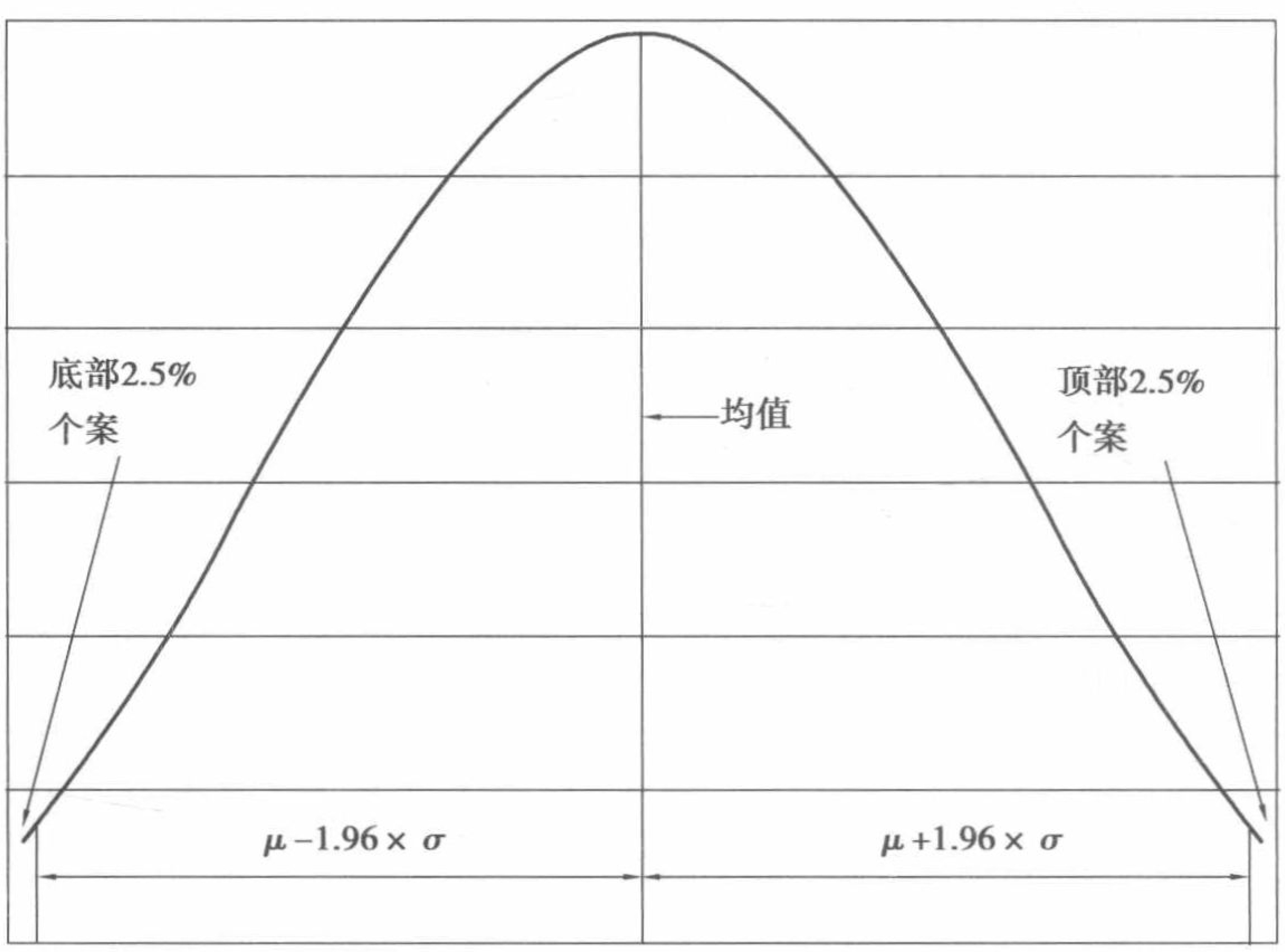



图1 钟形曲线实例

## 发展和应用史

社会统计学家对正态或钟形曲线一直持有浓厚的兴趣,尤其是在1810年,皮埃尔·拉普拉斯证明了中心极限定理之后,情况更加如此。这一定理显示如何从一个非钟形分布产生一条钟形曲线,以及一组有各自的误差的样本如何通过它们的均值呈现

为一种对称和可预测的形式。拉普拉斯的工作是建立在卡尔·弗里德里克·高斯工作的基础上的,而正态分布正是以高斯的名字命名的(故而有时正态曲线也称为高斯分布)。使用算术均值作为一种检查重复观察的值,并提供"真"值的天文学观察也在经验上显示,其他的一些观察也呈一种钟形曲线状,聚集在均值周围。因此钟形曲线也表示一组聚集在真值周围的误差。均值作为一种代表某种唯一的真值的思想,首先由阿道夫·凯特勒在他关于"平均人"的作品中提出来(参见平均数【Average】)。凯特勒也开发了一些(错误的)思想。他认为正态分布是一种社会和生物各个方面自然趋向的形式,而相反地,他假定在发现钟形曲线的地方,创造这一曲线的总体可认为是同质的。诚如施蒂格勒(Stigler,1999)指出的那样,在进入20世纪之后,认为钟形分布具有某些天然的合理性和诱惑的倾向已风光不再,然而,虽然在自然中某些分布近似一条钟形的曲线(例如成年人的高度),这样的曲线显示出围绕均值的聚集和向两侧的拖尾。但自然中的大多数分布却不是这样的曲线,而完美的对称分布几乎是找不到的。尽管如此,基于钟形曲线和它的性质的理论发展,特别是高斯分布在理论上的发展,仍然是社会科学统计推论的关键。

钟形曲线视觉和概念上的吸引力使它继续成为某些人捕捉真理和超社会的信仰的标记,尽管这只是借鉴它的统计推论的性质而已。而理查德·赫恩斯坦和查尔斯·默里(Herrnstein & Murray,1994)在他们多少有些声名狼藉的、有关美国白人和非裔美国人的智力分布的讨论中,使用"钟形曲线"一词作为他们的书名便是一例。

——Lucinda Platt

(沈崇麟译校)

## 参考文献


Aggresti, A., & Finlay, B. (1997). *Statistical methods for the social sciences* (3rd ed.) Upper Saddle River, NJ: Prentice Hall.

Herrnstein, R. J., & Murray, C. (1994). *The bell curve: Intelligence and class structure in American life*. New York: Free Press.

Marsh, C. (1988). *Exploring data: An introduction to data analysis for social scientists*. Cambridge, UK: Polity.

Stigler, S. M. (1999). *Statistics on the table: The history of statistical concepts and methods*. Cambridge, MA: Harvard University Press.

Yule, G. U., & Kendall, M. G. (1950). *An introduction to the theory of statistics* (14th ed.). London: Charles Griffin.


# 伯努利(Bernoulli)

伯努利,全名约翰·伯努利(John Bernoulli),瑞士数学家,以他的名字命名的概率【Probability】性质称为伯努利过程(Bernoulli process)。该过程用以描述有两个结果的事件的概率,例如掷一枚硬币。伯努利过程的 $n$ 次试验中 $k$ 次成功的二项【Binomial】概率由下式给出:

$$P(E) = \binom{n}{k} p^k (1-p)^{n-k}$$

式中的事件 $E$,是在 $n$ 次抛掷一枚公正的硬币中恰好有 $k$ 次面朝上的结果。

——Tim Futing Liao

(沈崇麟译校)

## 最佳线性无偏估计量(Best Linear Unbiased Estimator)

线性无偏估计量(BLUE)一词来自线性估计情形中,对无偏和高效估计一般概念的应用。在统计和经济研究中,我们几乎很少在总体中进行研究。在一般情况下,我们会有从一个总体中抽取的一个或若干个样本。我们的任务是,以作为估计量的样本统计量为依据,对参数的估计值做有意义的统计推论。在这里,估计量的统计性质是至关重要的,因为不具有上佳统计性质的估计量,我们是不可能对总体参数做可信的统计推论的。我们希望估计量是无偏的和高效的。换句话说,我们要求一个统计量产生的估计值的期望值能等于真的总体参数值。我们也要求由这一统计量产生的估计值的方差是所有无偏估计量中最小的。如果一个统计量满足了这两个条件,那么我们就称之为最佳无偏估计量(BUE)。BUE 估计量族,既可以是线性的,也可以是非线性的。如果 BUE 估计量来自一个线性函数,那么,我们就把它叫作最佳线性无偏估计量(BLUE)。

若要更为深入地了解 BLUE 一词的含义,我们则要深入详细地了解"估计量""无偏""最佳"和"线性"这些词的含义。

### 估计量

估计量是一种用样本统计值求总体参数值的决策法则。对于任何总体参数而言,无论它是总体平均数、离差,还是变量之间的关联量度【Measure of Association】,总是存在若干可能的估计量。好的估计量是那些平均而言,有着高度确定性的、接近总体参数的估计量。

### 无偏性

所谓无偏性是指,我们要求在反复进行的抽样中,由估计量产生的估计值的样本期望值等于真的总体值。例如,假设为了求总体回归函数$y_i=\beta_0+\beta_1x_i+\varepsilon_i$,对一个样本的双变量的线性回归【Linear Regression】$\hat{y}_i=\hat{\beta}_0+\hat{\beta}_1x_i+\varepsilon_i$ 进行估计、无偏性要求$\hat{\beta}_0$和$\hat{\beta}_1$的期望值等于总体中 $\beta_0$ 和 $\beta_1$ 的实际值。用数学语言可表达为$E[\hat{\beta}_0]=\beta$ 和 $E[\hat{\beta}_1]=\beta_1$。从直觉上讲,所谓无偏性,就是就平均而言,样本估计值将会产生参数的真值。不言而喻,这意味着我们所使用的估计法则,就平均而言,它产生的估计值,将会产生真的总体参数值。

### 效率

最佳一词涉及效率的标准问题。高效意味着一个无偏估计量,与其他所有无偏估计量相比,具有最小的方差。例如,假设我们有一个从正态总体抽取的样本。我们可以用均值或中位数来估计总体的中点。二者都是中点的无偏估计量。然而,均值确实是更有效的估计量,在重复抽样中,样本中位数的方差为均值方差的 1.57 倍。效率标准背后的原理是,设法提高样本估计值落入接近总体参数真值的概率。在重复抽样时,具有最小方差的估计量产生的估计值接近真值的概率较高。

### 线性

所谓估计量是线性的,是指我们要求因

变量的预测值是估计参数的一个线性函数。注意,在这里我们感兴趣的是参数的线性,而非变量的线性。例如,假设我们要估计诸如$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 \sqrt{x_i} + \varepsilon_i$ 这样的样本回归【Regression】。在这个回归中,$\hat{y}_i$和 $x_i$ 之间并非线性,但是$\hat{y}_i$和$\sqrt{x_i}$之间却是线性的。如果我们考虑将$\sqrt{x_i}$而非 $x_i$ 作为我们的解释变量,那么线性的假定就得到了满足。然而这种做法并不能应用于所有的估计量。例如,假设我们估计一个像

$$\hat{y}_i = \hat{\beta}_0 + \frac{\sqrt{\hat{\beta}_1}}{\hat{\beta}_0} x_i + \varepsilon_i$$

这样的函数,这个函数就不是一个线性估计量。因而在求$\hat{\beta}_0$和$\hat{\beta}_1$的时候,需要使用非线性估计法。一种这样的估计量可以是 BUE,但却可能不是 BLUE。

## 高斯-马尔科夫定理

BLUE 隐含着满足了上面讨论的三个标准。从数学的观点看,BLUE 是特别方便的。统计学中一个著名的成果,是高斯-马尔科夫定理。该定理显示,在某些假定下,最小二乘【Least Squares】估计量是最佳无偏估计量。这意味着在一个估计问题是参数线性时,它总是可以用最小二乘估计量无偏且方差最小地来估计总体参数。这个结果在数学上是十分方便的,因为它不需要去证明最小二乘的统计性质。只要高斯-马尔科夫假定得到满足,它们便总是具有良好的统计性质。

## BLUE 的局限

尽管 BLUE 的性质是很吸引人的,但这并不意味着这一特定的估计量总是一个最佳的估计量。例如,如果干扰并非正态,而是由一种"肥尾"的分布产生的,便常常会因此而发生较大的干扰,这时线性无偏估计量一般都会比其他考虑到这一点的估计量有着更大的方差。有鉴于此,非线性的或稳健估计量也许会更有效。此外,线性估计量并不是在所有的估计场合都是适合的。例如,在估计总体分布的方差($\sigma_e^2$)时,二次估计量则更适合。

——B. Dan Wood

Sung Ho Park

(沈崇麟译校)

## 参考文献

Davidson, R., & MacKinnon, J. G. (1993). *Estimation and inference in econometrics*. New York: Oxford University Press.

Greene, W. (2003). *Econometric analysis* (5th ed.). Englewood Cliffs, NJ: Prentice Hall.

Gujarati, D. N. (1995). *Basic econometrics* (3rd ed.). New York: McGraw-Hill.

Judge, G. G., Hill, R. C., Griffiths, W. E., Lutkepohl, H., & Lee, T. (1988). *Introduction to the theory and practice of econometrics*. New York: Wiley.

# Beta($\beta$)

$\beta$这个希腊字母以两种不同的方式与回归【Regression】的斜率有涉:第一种是作为总体回归方程回归斜率的值的符号(与样本相对而言);第二种是作为标准化斜率的符号(表示标准差【Standard Deviation】单位的变化),有时它也叫作"beta 加权"。通常

可根据上下文来弄清楚 beta 究竟是指两种含义中的哪一种。

——Michael S. Lewis-Beck

（沈崇麟译校）

＊也可参见标准化回归系数【Standardized Regression Coefficient】。

## 组间差(Between-Group Differences)

如果在两个变量之间有着共变,这就是说,如果一个变量的某些值常常与另一个变量的某些值相伴相随,那么两个变量之间便存在着一种关系。如果具有较高教育程度的人常常比教育程度较低的人有较高的收入,那么在教育和收入之间就存在着一种关系。

为了研究可能的关系,我们可根据自变量上的观察值来建组,并以这样的方法将所有在自变量上具有相同的值的元素都归入同一个组。这个自变量可以是定类变量【Nominal Variable】,如性别,这样所有的女性都在同一个组,而所有的男性则都在另一个组。接着,我们在因变量上对这两个变量进行比较,譬如说,对收入进行比较。如果我们发现两个组是有差别的,那么我们就说二者之间存在着某种关系。

自变量也可以是定量变量【Quantitative Variable】,如接受在校教育的年限。这时,任何接受过 8 年在校教育的人都属于同一个组,9 年的同属下一个组,以此类推。在这里我们再一次用收入作为因变量,比较不同组之间的收入。

一种进行组间比较的方法是先计算概括性统计量,如均值、中位数,然后跨组进行均值的比较。我们可能更乐于比较不同组的中位数,或者先创建诸如盒形图【Boxplot】这样的统计图,再对它们作比较。如果我们使用的是均值,那么这个问题就变成:我们应该怎么来比较它们? 此外,均值差与因变量值的散布相比较又有多大?

在自变量为定类变量时,使用最为普遍的比较组间均值的方法是方差分析【Analysis of Variance】,而在自变量为定量变量时,使用最多的则是回归分析。在方差分析中,我们将两种性别的收入均值与总均值相比较。在回归【Regression】分析中,我们求每一教育年预测的收入值,将它们与总收入相比较。因为这些点位于一条线上,我们无须问这些点是否有差别。相反,如果回归线通过这些点,且有一个非零的斜率,是否就意味着预测的收入是有差别的? 如果我们进行一个教育-收入数据的方差分析,我们可能会发现一个较小的残差平方和【Residual Sum of Squares】和一个比回归分析更好的拟合。但是方差分析使用了比相应的回归模型更多的自由度【Degrees of Freedom】,因而一般回归分析有一个较小的 $p$ 值【$p$ Value】。

在方差分析中,我们将组均值与总均值作比较,如果我们通过计算组间的平方和,将组的大小也考虑在内,组的变化越大,平方和就会变得越大。这一和数也叫作组间平方和【Between-Group Sum of Squares】。假定均值是有差别的,因为因变量受一个或多个自变量的影响。如果性别是一个自变量,那么这一和就可以标以性别平方和或模型平方和。

类似的数量在回归分析中,在大多数情况下都只叫作回归平方和。这一数量表示回归线的陡峭程度和观察值在自变量上的散布程度。

——Gudmund R. Iversen

（沈崇麟译校）

## 参考文献

Cobb, G. W. (1997). *Introduction to design and analysis of experiments.* New York: Springer-Verlag.

Iversen, G. R., & Norpoth, H. (1987). *Analysis of variance* (Sage University Paper Series on Quantitative Applications in the Social Sciences, series 07-001, 2nd ed.). Newbury Park, CA: Sage.

Lewis-Beck, M. S. (1980). *Applied regression: An introduction* (Sage University Paper Series on Quantitative Applications in the Social Sciences, series 07-022). Beverly Hills, CA: Sage.

# 组间平方和(Between-Group Sum of Squares)

从原则上讲,两个变量之间的关系的研究,首先要根据一个或多个自变量将数据分组。在每个组内,我们经常要计算自变量的概括性量度,如均值,然后再对这些均值进行比较。

如果均值不同,我们确定自变量和因变量之间存在某种关系。就两个组而言,比较是相对容易的。我们只要将一个均值减去另一个均值,看看它们是否有差别。但在有两个以上的分组时,均值的比较就变得相对复杂。我们既要了解各个组的均值的差别有多大,也要了解这些组有多大。

就一个有 $k$ 个值的定类变量【Nominal Variable】而言,我们可以用方差分析【Analysis of Variance】来确定这些组的差别有多大。在这样一种分析中,由于数学和历史的原因,我们计算如下的和数:

$$n_1(\bar{y}_1-\bar{y})^2+n_2(\bar{y}_2-\bar{y})^2+\cdots+n_k(\bar{y}_k-\bar{y})^2$$

式中带下标的 $n$ 是组内的观察数,而带下标上方有横杠的 $y$ 是 $k$ 个组的均值,而那个没有下标的均值就是因变量的总均值,即组间平方和。有时它被缩写为 BSS(between sum of squares),或模型平方和(model sum of squares)。很清楚,和的数量取决于组均值的差别的大小和这些组的大小。就若干个自变量而言,每一个变量都有组间平方和,通常它们都以变量名来表示。

组间平方和这一数量本身并没有告诉我们太多的东西。但是当它被总平方和除的时候,这一分数将告诉我们,因变量的变差中为这一自变量解释的比例。此外,当它被自由度($k-1$)除的时候,分数就变成了 $F$ 值的分子。而 $F$ 值则是用来确定是否拒绝无组间差别的虚无假设 $p$ 值【$p$ Value】的。

在自变量是定量变量【Quantitative Variable】时,通常回归【Regression】分析都要优于方差分析。线性模型假设因变量的总体均值位于一条直线之上。与方差分析不同,我们不需要用均值相等的虚无假设【Null Hypothesis】,我们只需要看一看回归线的斜率就可以了。一条斜率为零的水平线就意味着等均值。

回归的平方和取的是组间平方和的位置。它的主要优点是它的自由度【Degress of Freedom】数取决于自变量的数目(简单回归为 1),而不是取决于组的数目。然而,如果我们也可以对同一数据进行方差分析,那么 BBS 一般都会小于回归分析的平方和,这意味着方差分析对数据的拟合要优于回归。

——Gudmund R. Iversen

(沈崇麟译校)

* 也可参见平方和【Sum of Squares】。

## 参考文献

Iversen, G. R., & Norpoth, H. (1987). *Analysis of variance* (Sage University Paper Series on Quantitative Applications in the Social Sciences, series 07-001, 2nd ed.). Newbury Park, CA: Sage.

Lewis-Beck, M. S. (1980). *Applied regression: An introduction* (Sage University Paper Series on Quantitative Applications in the Social Sciences, series 07-022). Beverly Hills, CA: Sage.

# 偏倚(Bias)

在日常语言中,偏倚一词往往含有倾向(predisposition)、偏见(prejudice)或真相畸变(distortion)的意思。在统计学中,我们使用偏倚这个词的时候,虽然其一般含义与日常语言基本相同,但一般都与(样本)统计量【Statistic】(也称为估计量【Estimator】)和总体真值(参数【Parameter】)之间的关系有关。

试图估计一个统计量的期望值【Expected Value】和参数值之间的差的统计量叫作偏倚。为了对它的意义有全面的了解,我们可以像下面这样来做。

令 $X$ 为一个有某种概率密度函数【Probability Density Function, PDF】的随机变量【Random Variable】,它有如下的参数:(总体)均值 $\mu_x$ 和(总体)方差 $\sigma_x^2$。假设我们从这一总体抽取了一个有 $n$ 个观察的随机样本【Random Sample】。令

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

为样本均值。假设我们用 $\bar{X}$ 这一样本统计量作为总体均值 $\mu_x$ 的估计量。

如果 $\bar{X}$ 的期望值等于 $\mu_x$,也就是说 $E(\bar{X})=\mu_x$,那么我们就可以说,$\bar{X}$ 是 $\mu_x$ 的一个无偏【Unbiased】估计量。如果这个等式不成立,那么 $\bar{X}$ 就是 $\mu_x$ 的一个有偏估计量。这就是说,

$$偏倚 = E(\bar{X}) - \mu_x$$

不言而喻,如果样本均值的期望值等于真均值,该式就等于零。偏倚可以是正的,也可以是负的,它取决于 $E(\bar{X})$ 大于 $\mu_x$ 还是小于 $\mu_x$。

换句话说,如果我们重复抽取 $X$ 的 $n$ 个观察样本,这些重复抽取的样本均值的平均值应该等于 $\mu_x$ 的真均值。但是,请注意,我们无法保证任何来自一个样本的单个的 $\bar{X}$ 值必定等于真的总体均值。

为了确认样本均值的期望值是否等于真均值,我们可以像下面这样来做:

$$\begin{aligned} E(\bar{X}) &= E\left[\frac{1}{n}(x_1 + x_2 + \cdots + x_n)\right] \\ &= \frac{1}{n}[Ex_1 + Ex_2 + \cdots + Ex_n] \\ &= \frac{1}{n}[n.\mu_x] = \mu_x \end{aligned}$$

因为每个 $X$ 都有同样的均值 $\mu_x$。因为 $E(\bar{X})=\mu_x$,所以我们就可以说样本均值是总体均值的一个无偏估计量。

偏倚可以从几何学上用图 1 来阐述。

为了对偏倚作进一步的阐述,我们来看一下总体方差 $\sigma_x^2$ 的两个估计量,即

$$S_x^{*2} = \sum \frac{(X_i - \bar{X})^2}{n}$$

和

$$S_x^2 = \sum \frac{(X_i - \bar{X})^2}{n-1}$$

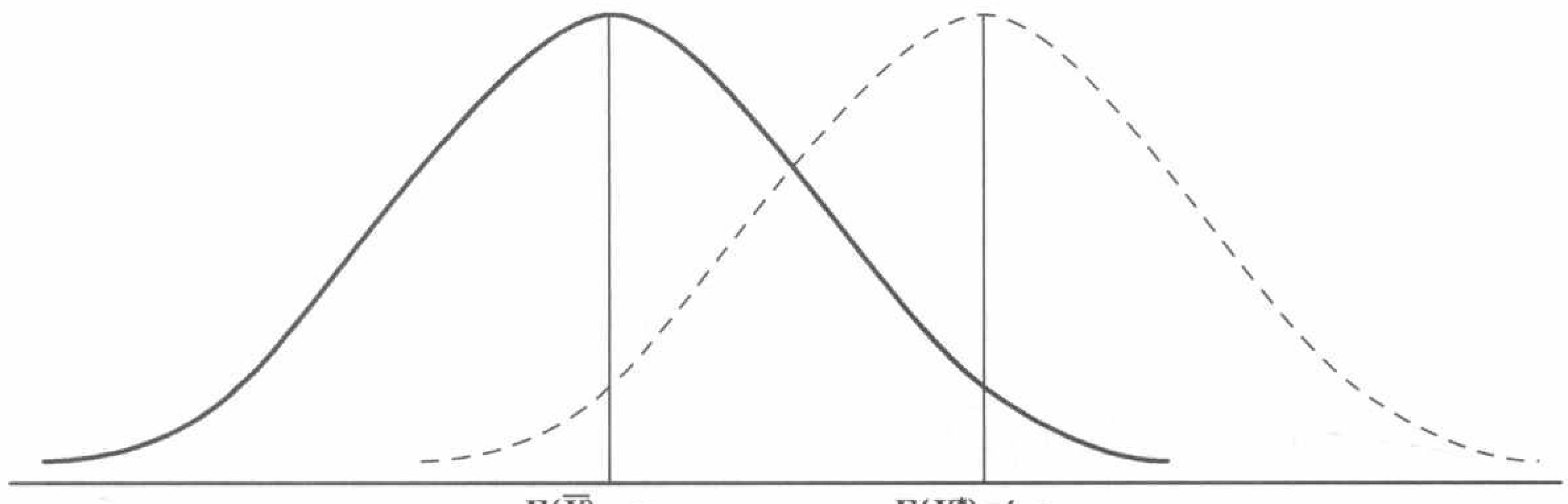



图 1 总体均值 $\mu_x$ 的有偏($x^*$)和无偏 $\bar{X}$

它们都是样本方差。这两个表达式的不同在于前者有 $n$ 个自由度【Degrees of Freedom】($df$),而后者则有($n-1$)个自由度。可以证明

$$E(S_x^{*2}) = \sigma_x^2\left(1 - \frac{1}{n}\right)$$

但是 $E(S_x^2) = \sigma_x^2$

因为前一个估计量的期望值不等于真的 $\sigma_x^2$,所以这是一个真方差的有偏估计量。

无偏性是我们希望的一个估计量应该满足的性质之一。它的基本思想直观而又简单。如果一个估计量被反复地使用,那么平均而言,它应当等于它的真值。如果这一点没有发生,那么它将系统地扭曲真值。

无偏性是一种比较小的或有限定的样本性质。一个估计量可以在小样本中不是无偏的,但在大样本中它却可以是无偏的。这就是说,随着样本量 $n$ 无限地上升,估计量的期望值或均值将接近它的真值。在这样的情形中,估计量叫作近似无偏的。这时情况很清楚,我们有估计量 $S_x^{*2}$,因为

$$\lim_{n\to\infty} E(S_x^{*2}) = \sigma_x^2$$

因此,$S_x^{*2}$ 是近似无偏的。在统计学中,我们会遇到各种各样的情况,这时一个在小样本中是有偏的估计量,随着样本量的无限增加,偏倚会随之消失。

迄今为止,我们给出的示例都只涉及一个随机变量。然而,偏倚这一概念非常容易被推广到多变量的情形。例如,让我们来看一下下面的多元回归:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \cdots + \beta_k X_{ki} + u_i$$

式中的 $Y$ 是因变量,带下标的 $X$ 是解释变量,而 $u_i$ 则是随机的【Stochastic】误差项。用矩阵记号,前面的多元回归可以被写作

$$\mathbf{y} = \beta\mathbf{X} + \mathbf{u}$$

式中的 $\mathbf{y}$ 是一个因变量上的观察值的($n\times1$)的向量,而 $\mathbf{X}$ 则是在 $k$ 个解释变量(包括截距)上的($n\times k$)的数据矩阵,而 $\mathbf{u}$ 则是($n\times1$)的随机误差项向量。

用普通最小二乘【Ordinary Least Squares, OLS】,Beta($\beta$)系数的估计量可由下式

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

求得。式中的 $\hat{\beta}$ 是真 $\beta$ 系数的估计量的($k\times1$)的向量。

在经典线性回归【Linear Regression】模型假定下(经典线性回归模型假定解释变量的值在重复抽样中是固定的,误差项是不相关的,且误差的方差是一个常数或同方差【Homoskedastic】),我们可以证明 $\hat{\beta}$ 是真 $\beta$ 的一个无偏估计量。证明过程如下:

$$\begin{aligned} E(\hat{\beta}) &= E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}] \\ &= \mathrm{E}[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\beta + \mathbf{u})] \end{aligned}$$

$$= \beta + E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}]$$
$$= \beta$$

这里使用了经典的线性回归模型假定(注意解释变量是非随机的)。

正如这些表达式显示的,估计的 $\beta$ 向量等于相应的真 $\beta$ 向量。换言之,每个 $\hat{\beta}_i$ 都是真 $\beta_i$ 的无偏估计量。

——Damodar N. Gujarati

(沈崇麟译校)

## 参考文献


Gujarati, D. (2002). *Basic econometrics* (4th ed.). New York: McGraw-Hill.


# BIC

参见拟合优度量度【Goodness-of-Fit Measures】。

# 双峰(Bimodal)

这是一种有着两个单独的巅峰或模式【Modes】的统计分布【Distribution】特性。与之相反,正态分布【Normal Distribution】则是一种单峰分布【Unimodal Distribution】。

——Tim Futing Liao

(沈崇麟译校)

# 二分(Binary)

顾名思义,二分的意思是二,在社会科学中它至少有三种含义。与十进位系统不同,二进位系统(binary system)只有两位数字(0 和 1),它是现代计算机的运算基础。二分变量(binary variable)测量一个只有两个类别的数量。在将社会现象分类的时候,社会科学家常常会进行二元思维(binary thinking),例如,公共的和私人的、礼俗社会和法理社会,以及宏观和微观等。

——Tim Futing Liao

(沈崇麟译校)

* 也可参见属性【Attribute】、二分变量【Dichotomous Variables】。

# 二项分布(Binomial Distribution)

二项分布有有限个独立的样本观察,每一个观察都可以假定是两种结果中的一个。因此二项分布是一个二分变量的分布【Distribution】,例如男性与女性,一枚硬币的面朝上和面朝下,活苗和死苗。样本【Sample】比例或概率【Probability】$p$,即一个给定的观察有一个给定的结果,是用样本观察有该种结果的计数 $X$ 除以样本量 $n$ 来计

算的。一对二分状态的每一种状态的概率之和总是等于 1.0。

二项分布与一个弹球落入一个由钉子制成的三角形阵列的 Galton 板的顶点得到的分布相似:落入的弹球走向左边或右边的可能性均为 50%。这也是由二项展开式的公式$(p+q)^n$造成的分布。该式中 $p$ 和 $q$ 是在 $n$ 次试验中得到或没有得到一个特定的二分结果的概率。例如(参见图 1),如果我们要从一个共和党人和民主党人人数相等的总体中选四个人,那么二项扩展式便是

$$
\begin{aligned}
(p+q)^4 &= (1/2+1/2)^4 \\
&= 1p^4 + 4p^3q + 6p^2q^2 + 4pq^3 + 1q^4 \\
&= 1/16 + 4/16 + 6/16 + 4/16 + 1/16
\end{aligned}
$$

计算二项分布的概率公式是

$$p(r) = {}_nC_r p^r q^{n-r}$$

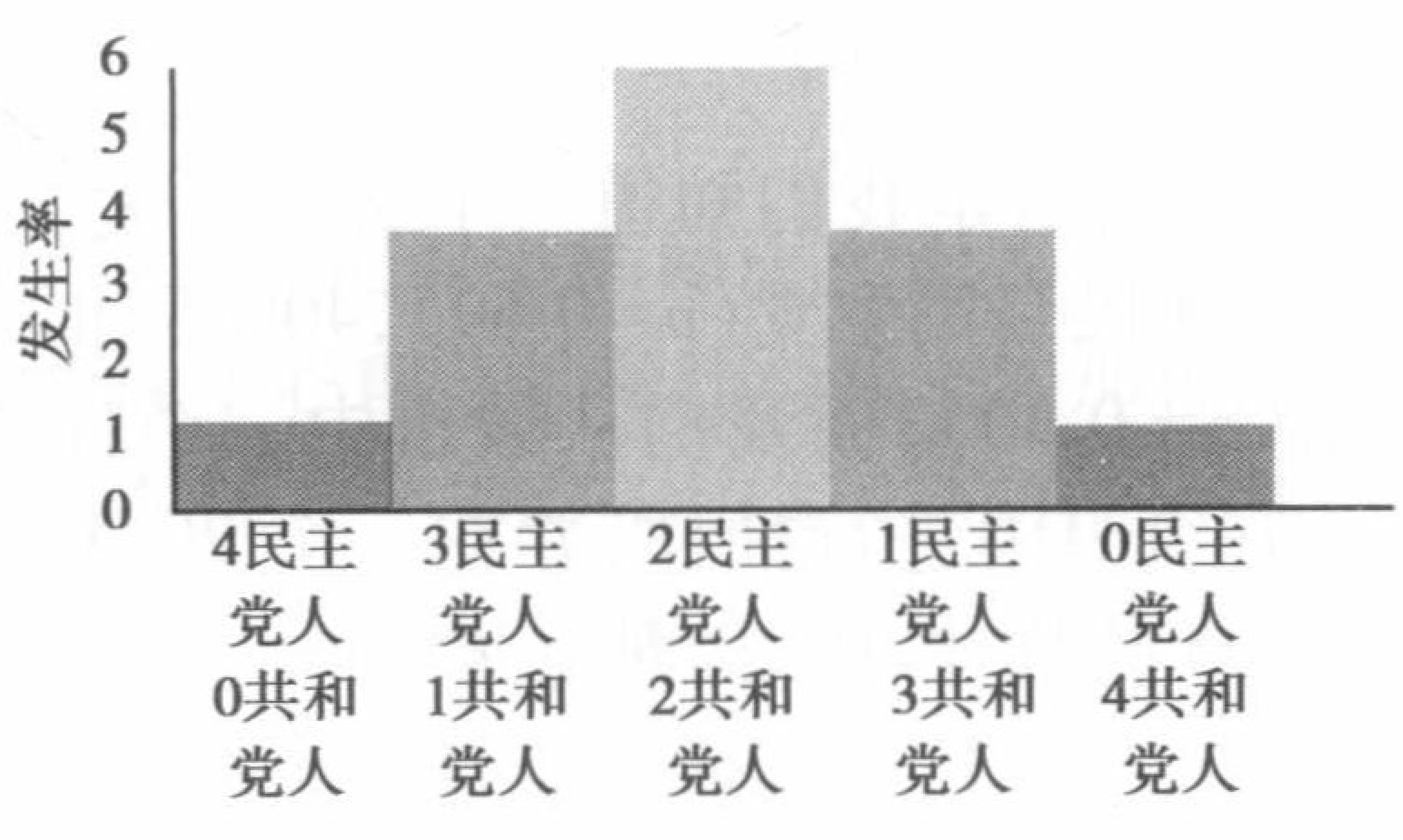



图 1 二项分布

式中的 $p(r)$ 是在二分变量【Dichotomous Variable】中的一个类别中得到 $r$ 个观察的概率,而得到另外一类的观察数则为$(n-r)$,如果样本量为 $n$ 的话;$p$ 是得到二分变量的第一种状态的概率,$q$ 是得到另一种状态的概率,而${}_nC_r$ 则是二分系数,它是每次从 $n$ 件事物中取 $r$ 件的组合数。在$p=q=0.5$ 时,二项分布将是对称的,否则就不是对称的。

例如,在抛掷硬币时,得到面朝上或面朝下的概率都是 0.5。抛掷三次硬币时,得到三个面朝上的概率是$(0.5)(0.5)(0.5)=0.125$。我们也可能得到一个面朝上、一个面朝上和一个面朝下,或者我们也可能得到一个面朝上、一个面朝下和一个面朝上。在三次抛掷中得到两次面朝上的组合数是${}_3C_2$,等于 $3!/[2!(3-2)!]$ 或 3。于是得到两次面朝上的概率就是

$$
\begin{aligned}
p(r) = {}_nC_r p^r q^{n-r} &= (3)(0.5^2)(0.5^{3-2}) \\
&= (3)(0.25)(0.5) \\
&= 0.375
\end{aligned}
$$

基于类似计算得到的二项分布表刊印在许多书刊中,或者它也可以用统计软件计算得到。$n=3$ 的二项分布表是

| $r$ | $p$ | | |
|---|---|---|---|
| | 0.1 | 0.5 | 0.9 |
| 0 | 0.729 | 0.125 | 0.001 |
| 1 | 0.243 | 0.375 | 0.027 |
| 2 | 0.027 | 0.375 | 0.243 |
| 3 | 0.001 | 0.125 | 0.729 |

因此,上表显示了 $p=0.1$,0.5 或 0.9 时,在三次试验中得到 $r=0,1,2$ 或 3 次面朝上或其他二分状态的概率。实际的二项分布表的列一般都是以 $p$ 的 0.1 的增量递增的,即 $p=0.1,0.2,0.3,0.4,0.5,\cdots,0.9$ 列出的,且同时还提供了从 $n=2$ 到 $n=20$ 的二项分布表。有些统计书刊既提供了一侧概率,也提供了双侧概率,前者适合用于“同样大/小”或“大于/小于”问题,而后者适合用于“相同或大大不同”的问题。

在样本量 $n$ 大,且发生的概率适中的时候,样本比例 $p$ 和计数 $X$ 将有一个接近正态分布【Normal Distribution】的分布。这一分布的均值将近似于 $np$,而它的方差将近似于 $np(1-p)$,它们是二项分布的均值和方

差。这些正态近似并不适用于小的 $n$ 值,因此一条普遍采用的经验法则是,只有在积 $np$ 大于或等于 10 和积 $np(1-p)$ 也大于或等于 10 时,才可以使用它们。如果积小于 10,那么就应该计算精确二项概率。统计软件一般不将正态近似用于二项概率,而是计算精确二项概率,即使对于大样本,也同样如此。

在样本量大,且发生的概率非常小或非常大的时候,二项分布呈泊松分布【Poisson Distribution】。在样本取自一个小的总体【Population】时,则要使用超几何分布而不是二项分布。

——G. David Garson

(沈崇麟译校)

## 参考文献

Blalock, H. M. (1960). *Social statistics.* New York: McGraw-Hill.

Von Collani, E., & Drager, K. (2001). *Binomial distribution handbook.* Basel, Switzerland: Birkhauser.

# 二项检验(Binomial Test)

二项检验是一种基于概率【Probability】法则进行的精确概率检验(exact probability test)。它用于在研究者有一个小样本【Sample】时,对单个二分变量的分布【Distribution】进行考察。它是一个非参数统计量,是一种用于单样本检验的样本比例和已知比例的差的检验。

进行二项检验要做四个假定。首先,它假定感兴趣的变量是一个二分变量【Dichotomous Variable】,它的两个值对所有的个案是互斥且穷尽的。第二,与所有的显著性检验【Significance Tests】一样,二项检验也假定数据是随机抽取的(参见随机抽样【Random Sampling】)。第三,它假定观察是独立抽取的。最后,二项检验假定任何观察的概率 $p$,有一个给定且跨所有的观察恒定的值。如果我们不能假定 $p$ 的无误差值,那么备择检验(费希尔检验、卡方检验【Chi-Square Test】和 $t$ 检验【$t$-Test】)则是我们所推崇的。二项检验是一种非参数检验,因而不用假定数据有正态分布【Normal Distribution】或任何其他特定的分布。

在二项检验中,如果样本量为 $n$,我们确定得到二类中的某一类观察为 $r$ 个,另一类的观察为$(n-r)$个。令 $p$=得到第一类的概率,$q=(1-p)$=得到另一类的概率。再令${}_nC_r$ 为一次在 $n$ 个事物中抽取 $r$ 个事物的组合数。二项概率的计算公式是

$$p(r) = {}_nC_r p^r q^{n-r}$$
$$= (n!\,p^r q^{n-r})/[r!\,(n-r)!\,]$$

例如,假定我们知道一个特定的城市,60%的居民是民主党($p=0.60$),40%是非民主党($q=0.40$),而我们从城市中一个特定的友好组织中进行抽样,并在一个 20 个人($n=20$)的样本中,发现有 15 个民主党人($r=15$)。此后,我们提出这样一个问题:我们样本中发现的 15∶5 分割比,是否显著不同于总体【Population】已知的分割比 60∶40?这个问题可以重新阐释为“得到一个样本分布概率的强度与观察的分布相同或更强的概率是多少?”答案是二项概率$p(15)$,$p(16)$,…,$p(20)$ 之和。例如,$p(15)=[(20!)(60^{15})(40^5)]/[(15!)(5!)]=0.074\,6$。由此可知,在 $n$ 相当大的时候,计算需要一台计算机。有些统计著作,也刊印了各种不同水平的 $n$,$r$ 和 $p$ 的单个或累积的二项概率表。

二项均值是 $np$,它是在 $n$ 个观察中的成功的期望数,二项标准差是 $npq$ 的平方根,它在 $p$ 接近 0 或 1 时趋于减小。

在 $n$ 大于25,$p$ 接近0.5和 $npq$ 至少为9的时候,二项分布近似于正态分布。在这样的情况中,正态曲线的 $z$ 检验【$z$-Test】也许可作为二项检验的近似使用。注意,只有在必须用手工进行计算时,我们才使用正态近似,因为各种计算机软件,如SPSS,总是可以为我们计算精确的二项检验。

——G. David Garson

(沈崇麟译校)

## 参考文献


Conover, W. J. (1998). *Practical nonparametric statistics* (3rd ed.). New York: Wiley.

Hollander, M.,& Wolfe, D. A.(1973).*Nonparametric statistical inference.* New York: Wiley.

Siegel, S. (1956). *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill.


# 传记叙事解释法(Biographic Narrative Interpretive Method,BNIM)

许多欧洲研究项目都以直接(例如,Chamberlayne, Rustin, & Wengraf, 2002; Rosenthal, 1998)或修改过的方式(例如,自由联想访谈)使用传记叙述解释这种方法来实施和分析传记叙述访谈。假定"叙述性表达"既能体现出有意识的关注,同时也可以体现无意识的文化、社会及个体的预设和过程,那么传记叙事解释法在研究方式上,便应当属于心理分析和社会传记的那种路数。

在每一个BNIM访谈中都有三个阶段(Wengraf, 2001, Chap.6)。第一阶段,访谈者只提出一个仔细设计的叙事问题(例如,"请告诉我你生活中的故事,对你个人重要的所有事件和经历,你可以从任何想要讲的地方讲起")。

在第二阶段,访谈者严格参照被访者提及的话题和使用词句的顺序,再让被访者提供更多的有关他们的故事。第三阶段则可以跟进形成一些非叙事性的问题。

以此方式获取的文字记录接下来会经过两次加工(Wengraf, 2001, Chap. 12)。在被访者解释事件、过着自己的日子和在访谈中诉说自己的故事时,研究策略重构那个"解释和表演"的被调查者的经历。这种研究策略要求分析者像被访者那样沿事件前行:对未来茫然(future-blind),时时刻刻永不停息,不知道接下来或以后会发生什么。

具体而言,首先,要对客观生活事件进行编年史确认。"客观生活事件"指的是那些能够通过官方记录,如学校、雇主或其他组织的记录查到的事件。每个事项(例如,16岁时考试失败)都在不考虑被访者现在解释的情况下,分别呈现给一个研究组,要求该研究组考虑这个事件在当时如何被经历,以及如果这个实验假设是正确的,接下来或者以后的生活中可能会发生什么(在遵从前述假设的前提下)。在研究者收集和记录这些内容之后,便会呈现下一个生活事件的事项:这意味着对先前的经验假设和后续假设的思考,继而开启新一轮的假设。研究者探寻有创造力的确认和批判性理解的过程,而这个过程则随着未来事件一个接一个地呈现,不断得到修正和精炼。

接下来文字记录被加工成"片段"。当叙述者、叙述话题或对话题的叙述方式发生变化时,一个新的片段就会开启。无独有偶,每个片段也都会依次呈递给研究组,而研究组则要试着想象这个访谈事件和行为是如何在访谈发生的那个时刻被经历,然后

再用后续的片段来进行修正和精炼。

在这两个序列中,研究者会探寻各自独立的结构性假设,并且只有之后才能发展出与这两个序列都相关的结构性假设。研究者也会将一个类似“对未来茫然”的程序,用于一个一字不差的文本中令人不解的片段(微观分析)。于是我们可就案例的动态提出这样的问题:“为什么人们以这种方式生活,却以那种方式讲述他们的故事?”

——Tom Wengraf

(郑晓娟译　沈崇麟校)

## 参考文献


Chamberlayne, P., Rustin, M., & Wengraf, T. (Eds.). (2002). *Biography and social exclusion in Europe: Experiences and life journeys.* Bristol, UK: Policy Press.

Rosenthal, G. (Ed.). (1998). *The Holocaust in three generations: Families of victims and persecutors of the Nazi regime.* London: Cassell.

Wengraf, T. (2001). *Qualitative research interviewing: Biographic-narrative and semi-structured method.* London: Sage.


# 传记法(Biographical Method)

参见生活史方法【Life History Method】。

# 双标图(Biplots)

双标图是用变量对受试者的数据矩阵进行各种分析的基本工具。在这样的分析中,研究者感兴趣的不仅是变量的结构,而且对个体之间的差异也感兴趣。令人惊异的是,在许多大型的统计软件中,它并非一个标准的工具。双标图还有几种变型,但我们在这里只介绍其中基本的一种。双标图可能是由塔克(Tucker, 1960)首先构建的,但加布里埃尔(Gabriel, 1971)在他的开创性工作中首先显示了这种方法的全部潜力。而目前有关这个方法的内容最为丰富的参考书则由高尔和汉德(Gower & Hand, 1996)所撰写。

一个标准的双标图展示为一个分解为有行坐标的 $I \times S$ 的矩阵 $\mathbf{A}=(a_{is})$ 和有列坐标的 $J \times S$ 的矩阵 $\mathbf{B}=(b_{js})$ 的积 $\mathbf{AB}'$ 的长方形矩阵 $\mathbf{X}$。其中下标 $S$ 表明坐标轴的数目,它可能是 2 或 3。用 $\hat{\mathbf{X}}$ 的这一分解,$\mathbf{X}$ 的一个二维近似这一矩阵的每个元素 $\hat{x}_{ij}$ 可写作

$$\hat{x}_{ij} = a_{i1}b_{j1} + a_{i2}b_{j2}$$

它是行向量$(a_{j1}, a_{j2})$和$(b_{j1}, b_{j2})$内(或标量)的积。通过在一个二维图中(有原点 0),将每一行示为有坐标$(a_{i1}, a_{i2})$的点 $a_i$ 和将每一列示为有坐标$(b_{j1}, b_{j2})$的点 $b_j$,便可得到一张双标图。一般我们把这些点分别叫作行标和列标。

例如,假设若干变量上的得分反映了那些初为人母的新妈妈育儿的有关情况。这些变量包括亲职压力(parenting stress)、抑郁(depression)和孩子认知刺激(cognitive stimulation of her child)等。这些数据的双标图将展示那些母亲,如她们在上述变量的得分所示,以相同的或不同的方式来育婴;

与此同时，双标图还显示，哪些变量有类似的模式（比如说，抑郁和亲职压力），或不同的模式（譬如抑郁和孩子认知刺激）。

在这样的情形下，变量一般都将被标准化，而变量的坐标则以一种变量和坐标之间相关的方式来量度（参见主成分分析【Principal Components Analysis】）。在本例中，变量有主坐标（principal coordinate），而应试者则有标准化的坐标，这就是说，他们的坐标轴有着 0 均值和单位长度。正因如此，变量通常显示为来自原点的向量。用向量的长度标示它的变异性，或显示为上面标出了原尺度大小的（双标图的）轴，即所谓的标刻度轴（calibrated axes）。应试者被标示为点，而变量上的应试者的估计分则从它的变量上的点的映射推出，如图 1 所示。因为在一个三维空间中，评估内积不是一件容易的事，所以最为常用的双标图只有二维和展示 $\hat{\mathbf{X}}$ 对数据矩阵 $\mathbf{X}$ 的最佳二维近似。

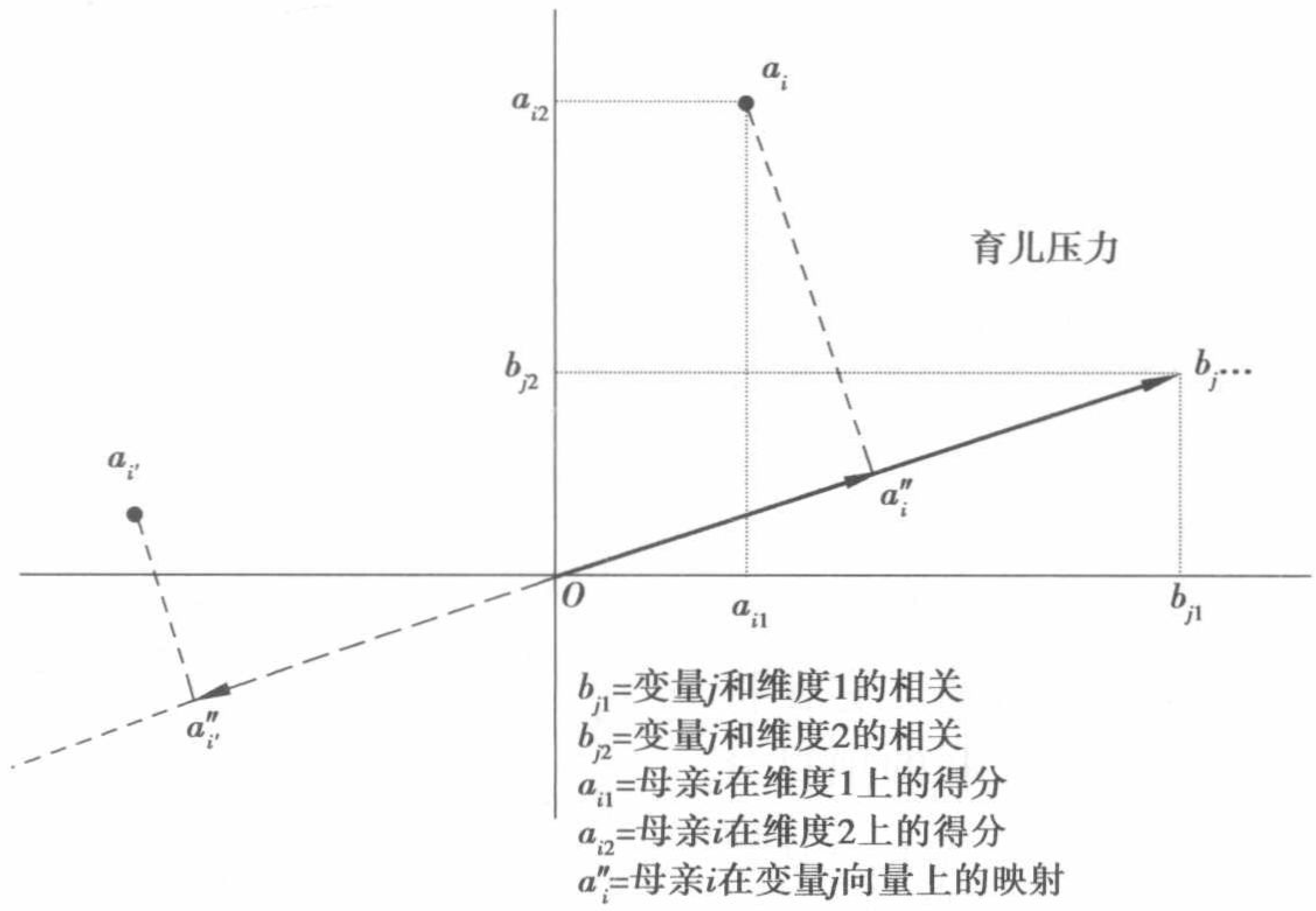



**图 1 双标图的主元素：列向量、行点、行坐标、列坐标和映射**

注：$a_i$ 和 $b_j$ 之间的内积是 $a_i''$ 和 $b_j$ 长度的积。

如果一个二维双标图是该数据矩阵的合理的近似，那么两个母亲在同一变量，比如说，亲职压力上的相对得分，便可以通过比较她们在亲职压力向量上的映射来判断。在图 1 中，这个比较是通过比较 $a_i''$ 和 $a_{i'}''$ 的长度实现的。如果一个母亲的得分在平均分之上，那么她的映射分将在变量正的一侧，反之则在负的那一侧。在她的映射分等于平均分时，它将与双标图的原点重合。两个相互间有一个锐角（钝角）的变量，有着正（负）相关。不相关则会产生正交的向量。所有这些陈述都是双标图对 $\mathbf{X}$ 中的数据足够好的近似条件。

在绘制双标图时，关键的问题在于物理的垂直和水平坐标轴应该有相同的物理尺度。这将保证行点映射在一个列向量时，它们将结束在正确的位置。不能做到这一点（有些计算机软件就没有做到这一点）将无法评估图中的内积。

双标图在除了自身是一种很有价值的方法之外，它在列联表的对应分析【Correspondence Analysis】和三模数据分析中也有着重要的作用。

——Pieter M. Kroonenberg

（沈崇麟译校）

## 参考文献

Gabriel, K. R. (1971). The biplot graphic display of matrices with application to principal component analysis. *Biometrika*, *58*, 453-467.

Gower, J. C., & Hand, D. J. (1996). *Biplots.* London: Chapman & Hall.

Tucker, L. R. (1960). Intra-individual and inter-individual multidimensionality. In H. Gulliksen & S. Messick (Eds.), *Psychometric scaling: Theory and applications* (pp. 155-167). New York: Wiley.

# 双极量表(Bipolar Scale)

双极量表是由相反的概念表示两端的量表，例如，冷/热、快/慢、不爱交际/爱交际、内向/外向。量表可以有 2~13 个刻度。

——Tim Futing Liao

(杨可译　沈崇麟审校)

# 双列相关(Biserial Correlation)

双列相关($r_{\text{bis}}$)是一种估计人为制造的二分变量($X$)和一个真的连续变量($Y$)之的关系强度的相关指数。两个变量都假定在它们的潜在总体中服从正态分布。它在将一个二分的得分项与总得分关联的项目分析中得到了广泛的使用。假定一个二分项的得分以这样的方式计——1 为正确答案，0 为不正确答案，那么双列相关则可这样计算：

$$r_{\text{bis}} = \frac{\bar{Y}_1 - \bar{Y}_0}{S_Y}\left(\frac{n_1 n_0}{u n_.^2}\right)$$

式中的 $\bar{Y}_1$ 是那些在 $X$ 上得分为 1 的人的均值；$\bar{Y}_0$ 是那些在 $X$ 上得分为 0 的人的均值；$S_y$ 是所有 $Y$ 得分的标准差；$n_1$ 是在 $X$ 上得分为 1 的观察数；$n_0$ 是在 $X$ 上得分为 0 的观察数；$n_. = n_1 + n_0$ 或总观察数；而 $u$ 则是在 $p_1 = n_1/n_.$ 的累积概率上的标准正态分布(即在 $X$ 上观察分为 1 的比例)的纵坐标(即垂直高度或概率密度)。

假定下面这组数据是从 30 个参加科学能力测试的考生中得来的。这些考生的测试总分和第 13 项测试的得分如下表所列：

| 考生 | 第 13 项得分($X$) | 总分($Y$) | 考生 | 第 13 项得分($X$) | 总分($Y$) | 考生 | 第 13 项得分($X$) | 总分($Y$) |
|---|---|---|---|---|---|---|---|---|
| 1 | 0 | 8 | 8 | 0 | 11 | 15 | 0 | 8 |
| 2 | 0 | 12 | 9 | 1 | 13 | 16 | 1 | 33 |
| 3 | 0 | 6 | 10 | 0 | 4 | 17 | 0 | 28 |
| 4 | 0 | 12 | 11 | 1 | 14 | 18 | 1 | 29 |
| 5 | 0 | 8 | 12 | 1 | 13 | 19 | 1 | 30 |
| 6 | 0 | 8 | 13 | 0 | 10 | 20 | 1 | 29 |
| 7 | 0 | 8 | 14 | 0 | 9 | 21 | 0 | 28 |

续表

| 考生 | 第 13 项得分(X) | 总分(Y) | 考生 | 第 13 项得分(X) | 总分(Y) | 考生 | 第 13 项得分(X) | 总分(Y) |
|---|---|---|---|---|---|---|---|---|
| 22 | 1 | 33 | 25 | 1 | 33 | 28 | 1 | 34 |
| 23 | 1 | 32 | 26 | 0 | 34 | 29 | 1 | 38 |
| 24 | 1 | 32 | 27 | 1 | 35 | 30 | 1 | 37 |

将总分作为 $Y$ 变量,而将第 13 项的分数作为 $X$ 变量。我们计算 $\bar{Y}_0 = 12.93$, $\bar{Y}_1 = 29$, $S_Y = 11.92$, $n_0 = 15$, $n_1 = 15$, $n. = 15+15 = 30$ 和 $p_1 = 15/30 = 0.50$。取 $p_1$ 的标准分布,我们确定纵坐标 $u$ 等于 0.398 9。将这些值代入上面的公式,得到

$$r_{\text{bis}} = \frac{29.00 - 12.93}{11.92}\left[\frac{(15)(15)}{(0.398\ 9)(30)^2}\right]$$
$$= (1.348\ 2)(0.626\ 7)$$
$$= 0.844\ 9 \approx 0.85$$

上面这个双列相关的值可解释为总分(连续变量)对第 13 项(人为制造的二分变量)上的正确与不正确答案的区分度。0.85 这个值说明总分与这一项的绩效有着很强的相关。如果 15 个没能完成这一测试的学生,他们的总测试分都是最低的,那么双列相关的值将等于 1.00(最大值)。相反,如果"通过"和"没有通过"第 13 项考试的,与总分的相配是随机的,那么双列相关的值便等于 0.00(绝对最小值)。

如果正态假定适用于 $X$ 和 $Y$ 的总体分布,那么双列相关便是一种积差相关(product moment correlation)估计,即皮尔逊的 $r$。它的理论值域从-1 到+1。在社会科学研究中,如果违反了正态假定或 $n$ 小于 15,那么根据经验数据计算的 $r_{\text{bis}}$ 的值就有可能超出理论的值域。因此只有在(a)未曾违反正态假定,(b)样本足够大(至少 100),(c)$p_1$的值与 0.5 无明显不同时,$r_{\text{bis}}$才可被看成皮尔逊的 $r$ 估计值。双列相关不可以用于诸如回归这样的二次分析。

——Chao-Ying Joanne Peng

(沈崇麟译校)

* 也可参见相关【Correlation】。

## 参考文献

Glass, G. V. & Hopkins, K. D. (1984). *Statistical methods in education and psychology* (2nd ed.). Englewood Cliffs, NJ: Prentice Hall.

# 双变量分析(Bivariate Analysis)

双变量分析是一种成对变量的统计分析。例如,这样的分析可以取交叉表【Cross-Tabulation】、散点图【Scatterplot】、相关【Correlation】、单因素方差分析【One-Way ANOVA】或简单回归【Regression】等形式。它经常是在数据【Data】单变量分析【Univariate Analysis】之后,但却在多变量分析之前采取的一个步骤。

——Tim Futing Liao

(沈崇麟译校)

## 双变量回归(Bivariate Regression)

参见简单相关(回归)【Simple Correlation (Regression)】。

## 分组设计(Block Design)

分组设计也称分块设计、区组设计。这是一种实验或准实验【Quasi-Experiment】设计。在这种设计中,为了控制观测结果或得分中非意愿的变差源,而将研究单位归并成类或"组"。非意愿的变差源是由那些对观测结果或得分有影响,但并非研究者主要感兴趣或关注的干扰变量(nuisance variables),也称为伴随变量(concomitant variables)引起的。

例如,在一个研究不同的治疗抑郁症方案的效应研究中,研究者也许希望控制抑郁症患者的严重程度。这样,在这个研究中,抑郁严重程度就是一个干扰变量。抑郁严重程度就是所谓的"组"。让我们进一步假定,研究者共有三种治疗方案和五个不同的抑郁严重程度。具体的设计方案如下表所示。

| | 治疗方案 1 | 治疗方案 2 | 治疗方案 3 | 行均值 |
|---|---|---|---|---|
| 组 1 = 最严重 | $Y_{11}$ | $Y_{12}$ | $Y_{13}$ | $\bar{Y}_{1.}$ |
| 组 2 | $Y_{21}$ | $Y_{22}$ | $Y_{23}$ | $\bar{Y}_{2.}$ |
| 组 3 | $Y_{31}$ | $Y_{32}$ | $Y_{33}$ | $\bar{Y}_{3.}$ |
| 组 4 | $Y_{41}$ | $Y_{42}$ | $Y_{43}$ | $\bar{Y}_{4.}$ |
| 组 5 = 最轻 | $Y_{51}$ | $Y_{52}$ | $Y_{53}$ | $\bar{Y}_{5.}$ |
| 列均值 | $Y_{.1}$ | $Y_{.2}$ | $Y_{.3}$ | 总净值 = $\bar{Y}_{..}$ |

假设分组设计的统计模型是

$$Y_{ij} = \mu + \alpha_i + \pi_i + \varepsilon_{ij}$$
$$(i = 1, \cdots, n; j = 1, \cdots, p)$$

式中 $Y_{ij}$——第 $i$ 组和第 $j$ 个治疗方案的得分。

$\mu$——观察总体的总平均,因此它是一个常数。

$\alpha_j$——第 $j$ 个治疗方案的治疗效应。在代数中,它等于总体均值($\mu_j$)与总均值($\mu$ 或 $\mu..$)的离差。第 $j$ 个方案的所有观察的依赖分(dependent score)是一个常数,如同固定效应模型那样,受跨全部治疗方案的所

有的 $a_j$ 之和为零的约束，或如同随机效应模型那样，在治疗总体中的所有 $a_j$ 之和为零的约束。

$\pi_i$——总体 $i$ 的组效应，它等于第 $i$ 个总体的均值（$\mu_i$）与总均值（$\mu..$）之离差。对所有观察在第 $i$ 组的依赖分它是一个常数，在潜在总体中服从正态分布。

$\varepsilon_{ij}$——与第 $i$ 组和第 $j$ 个治疗方案关联的随机误差效应。它是一个随机变量，在潜在的第 $i$ 组和第 $j$ 个治疗方案的潜在总体中服从正态分布，且独立于 $\pi_i$。

如果观察最初是随机分配给每个分组内的治疗方案的，这样的分组设计也叫随机化区块设计【Randomized-Blocks Design】。将观察分配给各个治疗方案的程序称为“分组”。有各种各样的将观察分组的方法，有时也可以根据干扰变量来对观察进行分组。这些分组方法的目标是：(a)使每一组含相等的干扰变量幅度段；(b)使每一组包含等比例的观察总体。大多数方法都试着去达到第二个目标。用上面的例子，研究者可以先按抑郁严重的程度（如按抑郁量表的得分）给病人排序。接着，研究者再随机地将第一组中得分最高的三个病人（最严重的抑郁症病人）随机地分配给三个不同的治疗方案。第二组由被认为抑郁程度比第一组较轻，但比其余病人重的三个病人组成。如此这般进行下去，直至把所有的组都建成为止。

如果一个以上的观察被分配到了分组和治疗方案的组合，那么这种设计就叫作广义分组设计（generalized block design）。作为一种选择，同一个病人（即一个观察）可以是一个组，并以随机顺序接受所有治疗。有些作者喜欢把分组设计的这一特定版本叫作重复测量【Repeated Measures】设计。

随机化的分组设计适用于满足普通的ANOVA假定（正确的模型、正态性、等方差和独立的随机误差）的实验，以及以下附加假定：

a.一个至少有两个水平和一个干扰变量的治疗因素，也可以有两个以上水平。

b.在每一组内，分析单位间的变异性应当小于不同组的分析单位间的变异性。

c.每个组内的观察应当随机地分配给所有的治疗水平。如果每个组只有一个在所有治疗水平进行观察的实验单位，那么对每个实验单位的治疗水平的呈现顺序应当是随机化和独立的（前提是治疗的性质允许）。

一般来讲，我们可有三种方法来控制观测结果中非意愿的变差源：

- 第一种，将干扰变量设为常数，例如只用中度抑郁的病人。
- 第二种，单位中的已知和未知的变差源在整个实验上是随机分布的，因而不会在特定的方向上只是影响一个或有限个治疗方案的假定下，将观察随机地分配给各个治疗方案。换言之，随机化的原则将自始至终地、反复不断地在同一个研究中对非意愿的变差源“予以关注”。
- 第三种，将干扰变量作为实验，如分组实验设计或协方差分析（ANCOVA）的因素之一。因此分组设计主要是为了更好地调查治疗变量（即 $\alpha_j$）的主效应，而不是与干扰变量（也包括分组）之间的交互效应。

——Chao-Ying Joanne Peng

（沈崇麟译校）

*也可参见分组设计【Block Design】、实验【Experiment】、多余参数【Nuisance Parameters】。

## 参考文献

Huck, S. (2000). *Reading statistics and research* (3rd

ed.). New York: Addison, Wesley, Longman.
Kirk, R. E. (1995). *Experimental design: Procedures for the behavioral sciences* (3rd ed.). Belmont, CA: Brooks/Cole.
Kirk, R. E. (1999). *Statistics: An introduction* (4th ed.). Orlando, FL: Harcourt Brace.
Maxwell, S. E., & Delaney, H. D. (1990). *Designing experiments and analyzing data: A model comparison perspective*. Belmont, CA: Wadsworth.

## BLUE

参见最佳线性无偏估计量【Best Linear Unbiased Estimator】。

## BMDP

BMDP 是一种统计分析软件,含有 40 种以上的统计分析程序。涵盖从简单的数据描述到复杂的多变量分析和时间序列分析等多个统计领域,不仅如此,BMDP 的输出结果是极具描述性的,且大部分都是不言自明的。

——Tim Futing Liao
(沈崇麟译校)

## Bonferroni 法(Bonferroni Technique)

Bonferroni 法是一种用于使族(family-wise)第一类错误【Type Ⅰ Error】率保持为一个事先选择的值的方法。这样的情形往往发生在对一个与显著性检验【Significance Testing】(或置信区间)有关的族进行处理的时候,有时我们也把这样的情况叫作多重比较【Multiple Comparisons】,或者更为广义地叫作“同步推论”。这种情形可能发生在同一个单位上进行的几次统计检验(例如在人们回答若干问卷题项时,或在同一个人身上计算若干变量的相关时),或在统计估计值,如均值或比例在一组假设上使用不止一次时(例如,一组计划的或事后进行的比较在统计上是不独立的)。虽然在这些检验是独立的时候,我们也可以使用 Bonferroni 法,但这样的情况十分罕见。

Bonferroni 法有几个其他同步推论法所不具备的长处,这使得它从各种备择方法中脱颖而出。在这些长处中有一种是它可以用于任何导致一个 $p$ 值的统计检验,因而使用起来十分方便。不仅如此,与其他的同步显著性检验相比较,它所浪费的验力也较少。

### 发展

Bonferroni 法的基础是邦费罗尼不等式,$\alpha_f \leqslant \alpha_1 + \alpha_2 + \cdots + \alpha_m$。因为 $\alpha_f$ 正是研究者想要控制的,所以选择几个 $\alpha_j$ 项,使 $\alpha_1 + \alpha_2 + \cdots + \alpha_m = \alpha_j$,这便会实现对 $\alpha_f$ 族第一类误差率的控制。通常,我们会将几个 $\alpha_j$ 设为彼此相等,且每个检验都在 $\alpha_t = \alpha_f / m$ 上进行。

因此 $\alpha_f$的值，最多为 $m\alpha_t$个。

## 应用

为了应用 Bonferroni 法，我们必须有一族（即组）$m$ 个（至少两个）显著性检验，并需定义两个显著性，即（$\alpha$）水平：每个检验的第一类误差率和族第一类误差率。每一检验的第一类误差率 $\alpha_t$ 是每一个统计检验的显著水平（即如果检验的 $p$ 值落在了 $\alpha_t$ 之下，我们就否定虚无假设）。族第一类误差率 $\alpha_f$ 是族中 $m$ 个检验中的一个或一个以上的显著度被宣布为第一类误差的概率（即那时虚无假设为真）。

在使用 Bonferroni 法时，我们需要对每个检验的第一类误差率进行校正，故而 $\alpha_t = \alpha_f/m$。在实际研究中，族第一类误差率通常都设为 0.05，所以，$\alpha_t$ 通常都设在0.05/$m$。

虽然各种分布（如 $t$、卡方分布）的那些不寻常的百分位数表都不难得到，但实际上使用 $p$ 值的 Bonferroni 法却是最容易使用的，因为它是统计软件的标准输出值之一。每个检验的 $p$ 值与计算得到的 $\alpha_t$ 相比较，如果 $p \leqslant \alpha_t$，那么我们就宣称统计结果是显著的，否则是不显著的。

## 实例

一个含四个题项的问卷发给了分性别的参与者，测验的目的在于了解男女在每个题项的分布上有什么不同。我们将使用基于四张双向表（性别与题项答案）的卡方检验【Chi-Square Test】。如果每个检验都在 $\alpha = 0.05$的水平上进行，它们当中至少有一个偶然达到显著水平的概率（假定所有的虚无假设为真）是大于 0.05 的，因为发生第一类误差的机会共有 4 个，而不只是一个。在这一例子中，我们有 $m = 4$ 个显著性检验，这样我们并不是要将每个检验的 $p$ 值与 0.05 作比较，而是要与 0.05/4 = 0.012 5 作比较。

如果要将 4 个实验组的均值中的每一个都与一个控制组的均值相比较，那么我们也同样可以使用 Bonferroni 法。检验可通过将 4 个检验（$t$ 或 $F$）每一个得到的统计值的 $p$ 值和 0.012 5 相比较来完成。虽然这些检验是相关的，因为它们合用一个均值（控制组的均值），但是族第一类误差率将被控制在不超过 0.05 上。

——William D. Schafer

（沈崇麟译校）

# 自助抽样推论法（Bootstrapping）

自助抽样推论法是一种计算机内含的集约非参数统计推论【Statistical Inference】法（computer intensive, nonparametric approach to statistical inference）。自助抽样推论法使用样本内部的变异性在经验上估计抽样分布，而不是对一个统计量的抽样分布【Sampling Distribution】做假定。做法是模仿原来的抽样方案，对样本进行许多许多次的回置再抽样。用这种估计的抽样分布进行统计推论，我们可有各种各样构建置信区间【Confidence Intervals】的方法。

推论统计的目的是根据从一个总体抽取的样本数据计算得到的统计量【Statistic】$\hat{\theta}$，对总体【Population】参数 $\theta$ 做概率【Probability】陈述。这样一种推论的核心是统计量的抽样分布。它是一个值域，它可以呈现为取自一个已知总体的随机样本和与这些值关联的概率。在社会科学家从研究生院学到的标准的参数推论统计（有无处不在的 $t$ 检验【$t$-Tests】和 $p$ 值【$p$

Values】)中,我们从数学分析中得到统计量的抽样分布信息。例如,中心极限定理【Central Limit Theorem】使我们有充分的理由相信,样本均值的抽样分布是正态的,有着总体均值的期望值和近似总体变量的标准差除以样本量的平方根的标准差【Standard Deviation】。然而,在有些情形中,既不存在诸如这样的参数统计理论,也不存在应用这种参数理论所需的假定。在遇到这些情况时,我们也许可以使用自助抽样推论法对总体参数做一种基于概率的推论。自助抽样推论法是一种广义的统计推论方法,实际上几乎可用于任何统计量。它的基本程序包含两个步骤:(a)通过重复抽样估计统计量的抽样分布;(b)使用估计的抽样分布构建推论总体参数的置信区间。

首先,将样本作为总体,对它进行一种蒙特卡洛模拟【Monte Carlo Simulation】,估计统计值的抽样分布。做法是从样本量为 $n$ 的原样本中,用回置法随机重复抽取大量样本量为 $n$ 的样本。回置抽样使得重复抽取的样本与原样本类似,但略有不同,因为原样本中的某个个案可能在任何给定的重复抽取的样本中,出现一次、一次以上或一次也没有。

重复抽样必须以模仿产生原样本的过程抽取。任何分层、加权【Weighting】、集聚、分段等抽取原样本时采用的方法,都必须在每一次重复抽样时同样采用。采用了这样的方式,那些注入原样本的随机变差,将会以类似的方式同样注入重复抽取的样本。根据复杂的随机样本【Random Sample】进行统计推论的能力,正是自助抽样推论法优于参数推论法之处。除了模仿原来的抽样程序之外,重复抽样只应该在统计模型的随机成分中进行。例如,这可以是简单描述统计的原始数据和调查样本,或回归【Regression】模型的误差项。

对每个重复抽取的样本,我们都要计算将要用于推论的样本统计量 $\hat{\theta}^*$。因为每个重复样本都稍微且随机地与另外的重复样本有所不同,所以这些 $\hat{\theta}^*$ 也会彼此稍微且随机地有所不同。自助抽样推论法主要判断的问题是,如果重复抽样的程序被用来导出在再抽样程序中被模仿的原样本,那么这些 $\hat{\theta}^*$ 相对频数分布【Frequency Distribution】应该是 $\hat{\theta}$ 抽样分布的一个估计值。

为了说明重复抽样的效果,让我们来考虑一下表 1 中的那个简单的例子。原样本是从一个标准正态分布抽取的简单随机样本。估计的均值和标准差与总体参数(分别为 0 和 1)略有不同,因为这是一个随机样本。请大家注意三个重复样本的若干情况。第一,在这些重复样本中不存在没有在原样本中出现的值,因为这些重复样本都产生自原样本。第二,因为重复抽样采用了回置法,所以某些原样本的值,在重复样本中出现了一次以上。第三,从重复样本估计的样本统计值(在本例中为均值和标准差),都接近但略有不同于原样本的这些值。这些均值(或标准差,或任何从这些重复抽样样本计算的统计值)的相对频数分布是总体参数的抽样分布的自助抽样推论法估计值。

**表 1　自助抽样推论法重复样本实例**

| 个案编号 | 原样本 [$N(0,1)$] | 重复样本#1 | 重复样本#2 | 重复样本#3 |
|---|---|---|---|---|
| 1 | 0.697 | −0.270 | −1.768 | −0.270 |
| 2 | −1.395 | 0.697 | −0.152 | −0.152 |
| 3 | 1.408 | −1.768 | −0.270 | −1.779 |
| 4 | 0.875 | 0.697 | −0.133 | 2.204 |
| 5 | −2.039 | −0.133 | −1.395 | 0.875 |
| 6 | −0.727 | 0.587 | 0.587 | −0.914 |
| 7 | −0.366 | −0.016 | −1.234 | −1.779 |

续表

| 个案编号 | 原样本 [$N(0,1)$] | 重复样本#1 | 重复样本#2 | 重复样本#3 |
|---|---|---|---|---|
| 8 | 2.204 | 0.179 | −0.152 | −2.039 |
| 9 | 0.179 | 0.714 | −1.395 | 2.204 |
| 10 | 0.261 | 0.714 | 1.099 | −0.366 |
| 11 | 1.099 | −0.097 | −1.121 | 0.875 |
| 12 | −0.787 | −2.039 | −0.787 | −0.457 |
| 13 | −0.097 | −1.768 | −0.016 | −1.121 |
| 14 | −1.779 | −0.101 | 0.739 | −0.016 |
| 15 | −0.152 | 1.099 | −1.395 | −0.27 |
| 16 | −1.768 | −0.727 | −1.415 | −0.914 |
| 17 | −0.016 | −1.121 | −0.097 | −0.860 |
| 18 | 0.587 | −0.097 | −0.101 | −0.914 |
| 19 | −0.270 | 2.204 | −1.779 | −0.457 |
| 20 | −0.101 | 0.875 | −1.121 | 0.697 |
| 21 | −1.415 | −0.016 | −0.101 | 0.179 |
| 22 | −0.860 | −0.727 | −0.914 | −0.366 |
| 23 | −1.234 | 1.408 | −2.039 | 0.875 |
| 24 | −0.457 | 2.204 | −0.366 | −1.395 |
| 25 | −0.133 | −1.779 | 2.204 | −1.234 |
| 26 | −1.583 | −1.415 | −0.016 | −1.121 |
| 27 | −0.914 | −0.860 | −0.457 | 1.408 |
| 28 | −1.121 | −0.860 | 2.204 | 0.261 |
| 29 | 0.739 | −1.121 | −0.133 | −1.583 |
| 30 | 0.714 | −0.101 | 0.697 | −2.039 |
| 均值 | −0.282 | −0.121 | −0.361 | −0.349 |
| 标准差 | 1.039 | 1.120 | 1.062 | 1.147 |

需要多少个这样的重复样本和 $\hat{\theta}^*$ 才能产生 $\hat{\theta}$ 切实可行的抽样分布的估计值呢?估计值是渐近无偏的,但是在什么条件下方差才能小到产生的推论精确度达到实用的程度。这个答案有两个组成部分:第一,就自由度而言,自助抽样推论法的抽样分布估计值的无偏性证明的渐近解,要求原样本的样本量在 30~50 个。第二,使估计的抽样分布充实和具体化的重复样本的数目似乎只需 1 000 个左右。对于高性能的计算机而言,如果研究者有能力写出恰当的循环算法【Algorithm】程序的话,这样的重复抽样和计算只需少量的时间和精力。

在用自助抽样推论法估计了 $\hat{\theta}$ 的抽样分布之后,自助抽样推论法统计推论的下一个步骤是,用这一个估计值构建置信区间。我们可以有几种方法来做到这一点。不过对究竟是哪一种方法得到的置信区间才最切实可用和在统计上最为正确这一问题,至今学界仍有争议。在 1980 和 1990 年代,有关自助抽样推论法的讨论的确推动并检验了这些方法,但这些都过于复杂,不宜在本文讨论。下面列出的参考文献中,有这些置信区间路数的详细论述和使用说明,有兴趣的读者可自行参阅。

## 发展

自助抽样推论法是近年来出现的一种统计方法,它最初见于布拉德利·埃弗龙发表在《统计学年鉴》(*Annals of Statistics*)的一篇题为《自助抽样推论法,折刀法再探》(*Bootstrap Methods: Another Look at the Jackknife*)的文章中。但是正如这篇文章所隐喻的那样,虽然埃弗龙普遍被人们认为是自助抽样推论法之父,但是确定样本内的变差,以把它用于统计推论的思想却并非新创。统计学家早就意识到参数统计的不足,并一直致力于克服它们。尽管参数统计推论在合适的时候是极为有力的,但是它能令

我们有足够的信心来使用它的范围是极其狭窄的,它只适用于某些特定的场合。罗纳德·弗希尔的精确检验是一种不需要参数假定进行推论的老方法,但是它要求完整地描述一个已知总体的全排列。显然,这对大多数现代的社会科学研究场合是不实际的。在1950和1960年代,曾经对几种使用各种样本统计量的变差来确定总样本的变差的方法进行过讨论。其中最为普遍的是折刀法,这种方法系统地从一个样本中,一个一个地去掉个案,并判断这种一个个的个案去除对统计量变差的影响。但直到高性能的个人计算机的横空出世,埃弗龙才能在这方面有重大的突破——使用无限个随机生成的子样本来估计统计量的抽样分布,于是自助抽样推论统计法应运而生。

## 应用

有两种情况是自助抽样推论法最有可能为社会科学家所用的:(a)在用一种没有很强的参数理论与之关联的统计量进行推论的时候;(b)在用一种有在某些条件下有参数理论,但这些条件却并不具备的统计量进行推论的时候。例如,自助抽样推论法可以用于第一种情形的有:从非直接的路径模型效应、特征值【Eigenvalues】、切换回归的切换点或两个中位数的差作推论等。随着计算机功效变得越来越强和社会科学家使用的统计方法变得越来越复杂,需要设计开发能非常精确地检验特定的假设的估计量,但却不用掌握这些估计量的任何抽样分布的性质,而使自助抽样推论法变得尤其重要。在遇有违反了参数检验的假定,而必须对参数统计检验的稳定性进行检验,则成为自助抽样推论法变得重要的第二个原因。例如,因为最大似然估计【Maximum Likelihood Estimation, MLE】的近似参数理论,似乎只有在样本量达到100时才站得住脚。标准的 $t$ 检验在样本较小,比如只有75的时候,就不太精确。自助抽样推论法的统计推论便可以在诸如这样的情形中使用,以了解这样的不精确性是否真正会成为一个问题。

——Christopher Z. Mooney

(沈崇麟译校)

## 参考文献

Chernick, M. R. (1999). *Bootstrap methods: A practioner's guide.* New York: Wiley-Interscience.

Davison, A. C., & Hinkley, D. V. (1997). *Bootstrap methods and their application.* Cambridge, UK: Cambridge University Press.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *Annals of Statistics*, *7*, 1-26.

Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap.* New York: Chapman & Hall.

Mooney, C. Z., & Duval, R. D. (1993). *Bootstrapping: A nonparametric approach to statistical inference.* Thousand Oaks, CA: Sage.

# 有限理性(Bounded Rationality)

有限理性是用来描述决策过程的术语。它不太像一个统一的理论,更像一种认识,即一种认为实际的决策与作为许多行为形式模型基础的理性概念不符的认识。有限理性这个术语是1950年代由赫伯特·西蒙提出的(Simon, 1957),在今天的研究中仍然是一个很活跃的领域(Rubenstein, 1998)。值得注意的是,有限理性并不是非理性。在有限理性的指引下,行动者仍然还是试图做出好的决策,只不过他们的决策过

程不能满足完全理性的所有要求而已。

理解有限理性的最佳方法也许就是将之与大多数行为形式模型所假设的完全的理性行动者作比较。如果信息可以得到的话,完全的理性行动者总是能做出最佳的决策。完全的理性行动者有清晰定义的偏好,以无偏的方式使用所有相关的信息,而且能够准确无误地解决棘手的最优化问题。尽管大多数采用完全理性假设的研究者都明白,这不是对现实中决策的完全准确的描述,但他们还是常常假定行动者的行为表现"好像"是完全理性的。

尽管如此,那些研究有限理性的学者还是找到了许多行动者的行为表现并非"好像"是完全理性的例子。实际上,行动者可能无法完成完全理性行为要求的复杂计算。例如,在一个无限重复的游戏中,为了计划行动者当今的行动,完全理性要求行动者将无数个周期考虑进未来,然而,这很有可能超出了组织或个人的计算能力,甚至即使是最有能力的组织和个人也办不到。在有限理性的指导下,行动者也还是试图为当今的最佳行动做出安排,但却无法进行完全理性的行动者所做的那种复杂计算。可能有限理性和无限理性最一般的区别应归功于西蒙,他认为,行动者会在完全理性的情况下进行最优化,即寻求一个问题的最佳解决方案,而在有限理性的情况下,行动者通常会采用满意策略,采用一个"足够接近"的解决方案。

学者们还曾将其他一些有限理性行为的例子记录在案,例如简化问题的倾向和框架效应。完全理性的行动者对逻辑上等价的备择选项和选择集合的说明并不关心。不过,特沃斯基和卡内曼(Tversky & Kahneman,1986)在一个著名的实验中证明了判定问题的框架会对个人的选择产生影响。实验的被试者被告知,美国会爆发让600人丧命的疾病,然后实验的被试者被分成两组。第一组被试者要在两个互斥的治疗方案(A和B)中选择一个,这两个方案可能会产生如下结果:

A.200人会得救。

B.600人得救的概率是1/3;一个人也不会得救的概率是2/3。

第二组被试者要在两个互斥的治疗方案(C和D)中选择一个,这两个方案可能产生如下结果:

C.400人会死亡。

D.一个人也不会死的概率是1/3;600人都会死的概率是2/3。

显然A等价于C,而B则等价于D。但是,第一组里72%的人选择了A方案,而第二组里78%的人选择了D方案。这个结果进一步证明,在实际的行为和完全理性的假设之间有不少差异,因此,许多决策问题中的行动者最好是被称作有限理性人。

——Garrett Glasgow

(杨可译 沈崇麟校)

## 参考文献

Rubenstein, A. (1998). *Modeling bounded rationality.* Cambridge: MIT Press.

Simon, H. A. (1957). *Models of man.* New York: Wiley.

Tversky, A., & Kahneman, D. (1986). Rational choice and the framing of decisions. *Journal of Business*, *59*, 251-278.

# 盒须图(Box-and-Whisker Plot)

参见盒形图【Boxplot】。

# Box-Jenkins 建模法(Box-Jenkins Modeling)

与 Box-Jenkins 这两个名字关联的研究方法主要用于等间隔的时间序列【Time-Series】数据(Box & Jenkins, 1976)。与确定性模型【Deterministic Model】不同, Box-Jenkins 模型将时间序列作为一个随机【Stochastic】过程,特别是作为一种自回归整合移动平均(ARIMA: autoregressive, integrated, moving average)的过程对待。Box-Jenkins 模型主要应用于预测一个时间序列的未来观察值,确定一个进行中的时间序列的干预效应(干预分析【Intervention Analysis】)和估计动态的投入产出的关系(传递函数分析)。

## ARIMA 模型

最简单的 ARIMA 模型假定一个时间序列的观察值($z_t$)是根据移动平均参数($\theta$),由只延续到下一个时间点的随机冲量($u_t$)产生的。这是一个一阶移动平均模型, MA(1),它有形式(忽略了参数的负号):

$$z_t = u_t - \theta u_{t-1} \tag{1}$$

这一过程只离开了纯随机性一小步,叫作"白噪声(white noise)"。今天的记忆($t$)只继承了昨天的新闻($t-1$),而不继承任何此前的新闻。即使如此,如果需要,一个移动平均模型仍然能轻而易举地扩展,以容纳一大组先前的随机冲量,比如 $u_{t-2}$或 $u_{t-3}$,但这时,MA 这一框架就会显得十分笨拙。有一种更为实用的处理累积的,但是已贴现的过去冲量的方法,那就是自回归【Autoregressive】模型法。它将把我们引向 ARIMA 的"AR"部分:

$$z_t = \phi z_{t-1} + u_t \tag{2}$$

在这一过程中,AR(1)累积的过去的($z_{t-1}$),不仅仅只是最近的随机冲量($u_{t-1}$),而是根据自回归参数($\phi$),继承了一个时期到下一个时期的随机冲量。但并非是全部时期。自回归过程要求在从时间 $t-1$ 向时间 $t$ 过渡时有某种泄漏量。只要 $\phi$ 在 1.0 以下,那么 AR(1)过程就满足了稳定性(常均值水平、常方差和观察值之间的常协方差)的要求。

当然在实际生活中,许多时间序列是不稳定的。它们展示某些趋势(长期上升或下降),或自由徘徊,就像随机游走【Random Walk】。这样的时间序列的观察值,在自回归和/或移动平均成分可以被估计之前,必须转换成稳定的观察值。标准的程序是做时间序列 $z_t$ 的差分,

$$\nabla z_t = z_t - z_{t-1} \tag{3}$$

并确定得到的系列($\nabla z_t$)是否满足了稳定的要求。如果满足,原系列($z_t$)便被认为在序 1(order 1)被整合了。这就是在 ARIMA 中"*I*"的含义。*I* 无非是一个时间序列在达到稳定之前必须被差分的次数的计数而已。对大多数不稳定系列而言,差分一次($I=1$)就足够了。

然后便可开始进入稳定的时间序列分析过程,先进行模型识别,然后进行参数估计和模型的诊断检查。为了确认稳定的时间序列的动态是自回归(AR)或移动平均(MA),还是二者的组合——再加上这些成分的次序,我们要不断地考察一系列自相关【Autocorrelations】系数和偏自相关系数,直至有足够的滞后数为止。任何 ARMA 过程都会在这些相关函数(ACF 和 PACF)留下自己的特征标志。然后我们通过迭代最大似然估计【Maximum Likelihood Estimation】程序,估计被认定的模型的参

数。这一模型对捕捉 ARMA 的时间序列动态是否精确,要取决于误差诊断。Ljung-Box 的 Q 统计量被广泛地用来进行白噪声残差的概括性检验。

## 预测

Box-Jenkins 模型被设计用来避免确定性函数的预测误区(外推时间趋势)。一个 ARIMA 预测发端于随机动态地移动一个时间序列。一旦模型参数被求得,就不再需要其他数据便可进行预测。预测自动地随模型的估计而来,它们先于被预测事件到来。因此 ARIMA 是真正事前的无条件的预测,它们包含了有关现象的动态性状的所有信息。例如一个美国总统选举的 AR(2) 模型,用 1860 到 1992 年的结果聚集进行估计,很早就能预测到 1996 年民主党的胜利(Norpoth, 1995)。

## 干预效应

有许多形式的行为和绩效是政府想通过新政的实施来加以改变的。无论这些结果是否经常被测量,Box-Jenkins 模型都能给出诸如这样的干预效应的估计值(Box & Tiao, 1975; Clarke, Norpoth, & Whiteley, 1998)。一个干预模型,比如说,一个旨在降低交通死亡事故的政策,其基本结构如下所示:

$$y_t = [\omega/(1-\delta B)]I_t + N_t \qquad (4)$$

式中 $y_t$——第某个月的死亡数;

$I_t$——干预变量(1 政策实施月,0 政策未实施月);

$\omega$——与干预变量关联的步长变化(step change);

$\delta$——步长变化的死亡率;

$B$——反向移位算子(诸如这种情况的例子,$BI_t = I_t - 1$);

$N_t$——按 ARIMA 过程建模的“噪声”。

这个模型既不需要干预的影响是持久的,也不需要干预终止影响同时消失。通过将一个时间序列过程按 ARIMA 过程来建模,我们可避免把宣称的干预的影响与其他可能的效应混杂。诸如这样的随机化使我们不会误将一种趋势当作干预的效应。

## 传递函数

与改变了时间序列过程的离散的干预不同,让我们来考虑一下另一种连续地影响着我们感兴趣的时间序列的上下波动的时间序列。这种影响不必是即时的,但可能需要一定时间充分成形。这种时间序列可用一个传递函数表达为如下的形式(Box & Jenkins, 1976):

$$y_t = [\omega/(1-\delta B)]x_t + e_t \qquad (5)$$

在这个模型中,$\omega$ 和 $\delta$ 描述了一种不断地在 $X$ 和 $Y$ 之间发生的影响。$X$ 移动时,$Y$ 也移动,而移动的量则由参数 $\omega$ 设定。而稍后,$X$ 的移动又将引起 $Y$ 根据参数 $\delta$ 作出进一步的反应,如此这般,连续进行。为了识别这种每个变量都高度自相关的变量动态模型,博克斯-詹金斯(Box & Jenkins, 1976)引进了下面这样的策略:先通过去掉 $X$ 的 ARIMA 动态,将 $X$ 系列“预白化”,然后相应地筛选 $Y$ 系列,并求得变换的时间序列的交叉相关。这种策略,如同一剂猛药,最终可能导致“将孩子连同洗澡水一起倒掉”。所以我们提醒读者,在使用这种策略时,务必倍加小心。

虽然 Box-Jenkins 模型涵盖了许多时间系列问题,但它毕竟不是一剂无所不能的灵丹妙药。这些模型的用户,可能会面临最终一无所得的风险。时间序列是一个

这样的领域，在这个领域里，近乎完美的相关可能在一刹那间化为随机性。差分一个随机漫步，连续的观察完全相关，而得到了白噪声，则相关为零。

——Helmut Norpoth

（沈崇麟译校）

## 参考文献


Box, G. E. P., & Jenkins, G. M. (1976). *Time series analysis: Forecasting and control.* San Francisco: Holden-Day.

Box, G. E. P., & Tiao, G. C. (1975). Intervention analysis with applications to economic and environmental problems. *Journal of the American Statistical Association*, *70*, 70-79.

Clarke, H. D., Norpoth, H., & Whiteley, P. F. (1998). It's about time: Modeling political and social dynamics. In E. Scarborough & E. Tanenbaum (Eds.), *Research strategies in the social sciences* (pp. 127-155). Oxford, UK: Oxford University Press.

Granger, C. W. J., & Newbold, P. (1986). *Forecasting economic time series* (2nd ed.). New York: Academic Press.

Mills, T. C. (1990). *Time series techniques for economists.* Cambridge, UK: Cambridge University Press.

Norpoth, H. (1995). Is Clinton doomed? An early forecast for 1996. *Political Science & Politics*, *27*, 201-206.


# 盒形图(Boxplot)

盒形图是由图基(Tukey, 1977)开发的一种用以概括一个分布的中点、范围和将它与其他一些分布或一些用另一个给定的定类变量分组的分布进行比较的方法。盒子的大小表示选定的变量中间的50%的值的范围，也即从第25分位到第75分位的范围，或者如大家可能已经知道的那样，四分位距【Interquartile Range】的范围。一条穿过这个盒子的线表示中位值。从盒子延伸出来的须子(whiskers)到达了位于最远两端的值，但异常值则被排除在外。这就是说，它们到达的无论是哪一端，最远的值都在盒子大小的1.5倍之内。我们对于须子长度的限制叫作栅栏(fences)。不过须子能达到的最远的值则不一定非要到达栅栏。至于究竟能否到达栅栏则要取决于分布的形状和离差的量。异常值是在栅栏之外的值，因为显著离群，因而可予以个别标示。极端值(extremes)是异常值的极端版本，它们落在盒子无论哪一端的盒子大小的3倍以外的地方。它们可以一一被记下和标示。可见盒形图是一种高度浓缩的图形形式，为我们提供大量的有关分布的重要信息，不仅如此，它对确认分布的极端和异常值更是特别有用。

图1给我们提供了一个比较1999年，英国成年男女周薪的盒形图实例。盒形图表明了薪水分布的形状。我们可以看到男女两性的图形都是向右偏斜的(在靠近0处聚集)。不过女性的中位数位于盒子较低的位置，说明女性有更大的低收入的群。此外，女性的整个分布位于较低的位置，这说明她们的薪水普遍偏低。图中有一定数目的奇异值(用个案号标示)，说明有为数不多的人有着很高的薪水。我们注意到，大多数女性的奇异值都位于男性的栅栏内，对此，我们颇感兴趣。

图2是一个集群条形图。它将图1用教育水平作了分解，但图中每个子类提供的信息与图1一样。它探索的问题是，观

察值在用性别和教育程度分类后，周薪是否仍然显示与图 1 同样的分布模式。结果显示，其他教育程度的女性的周薪分布不仅显得十分扁平，而且不存在异常值。图 2 还显示，跨 4 个组的四分位互差稳步变窄，这说明最高学历的男性的周薪的异质

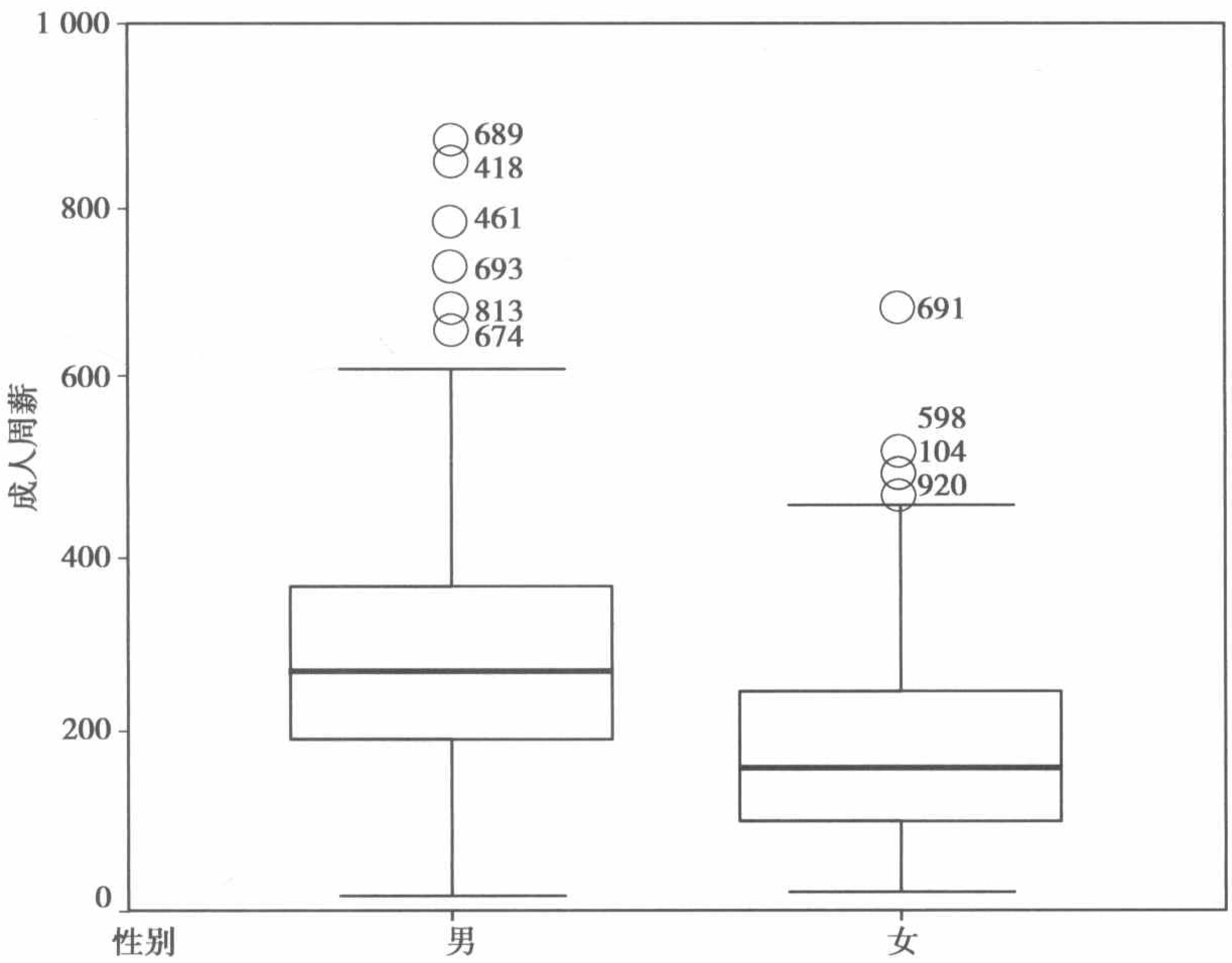



图 1　男女性周薪盒形图

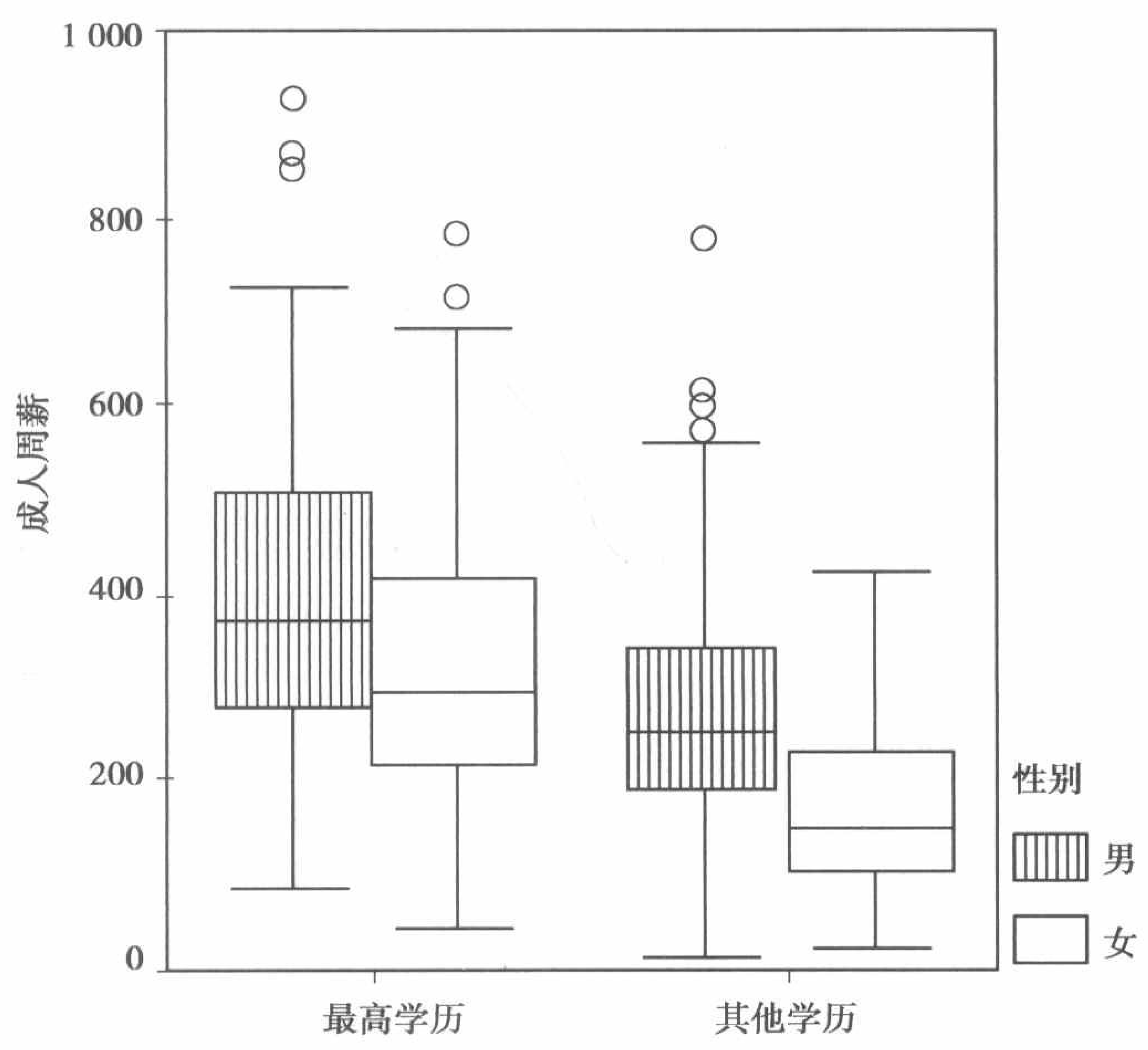



图 2　分性别和教育程度周薪盒形图，1999 年，英国

性高，而教育程度较低的女性的周薪最具同质性。这表明图形表述的长处是不仅能表达数据的基本信息，而且也能鼓励我们对它进行深入的探索。

——Lucinda Platt

（沈崇麟译校）

## 参考文献

Marsh, C. (1988). *Exploring data: An introduction to data analysis for social scientists.* Cambridge, UK: Polity.

Tukey, J. W. (1977). *Exploratory data analysis.* Reading, MA: Addison-Wesley.

# C

## CAIC

参见拟合优度量度【Goodness-of-Fit Measures】。

## 典型相关分析(Canonical Correlation Analysis)

最初对典型相关分析的基本思想予以清晰阐述的是拉西特(Russett,1969)。他分析了47个国家某些经济和政治特性之间的关联【Association】。他用了5个指标来测量经济的不平等性:农田分割(the division of farmland)、基尼系数【Gini Coefficient】、佃农百分比、人均国民生产总值(GNP)和农民百分比。拉西特用4个指标来测量政治不稳定性:领导不稳定性、群体内暴力程度(level of internal group violence)、内战爆发和民主制度稳定性。研究假设托克维尔是正确的:如果经济资源在公民中的分配是不均衡的,那么没有一个国家能在一个相当长的时期内保持政府的民主形式。换言之,国家的经济不平等性和政治不稳定性之间是明显关联的。

在这个研究问题中的两个理论概念是"经济不平等性"($X^*$)和"政治不稳定性"($Y^*$)。它们叫作有着期望高相关的典型变量,即所谓的典型相关(canonical correlation)。第一个典型相关变量$X^*$由$p=5$个指标的$X_1$到$X_5$测量,我们将考虑把$X^*$作为这些$X$变量的线性组合。以此类推,$Y^*$第二个典型变量是有$q=4$个指标的$Y_1$到$Y_4$的线性组合。在最为一般的情形中,$X$组含有$p$个变量,而$Y$组则含有$q$个变量,它可以用图1表示如下:

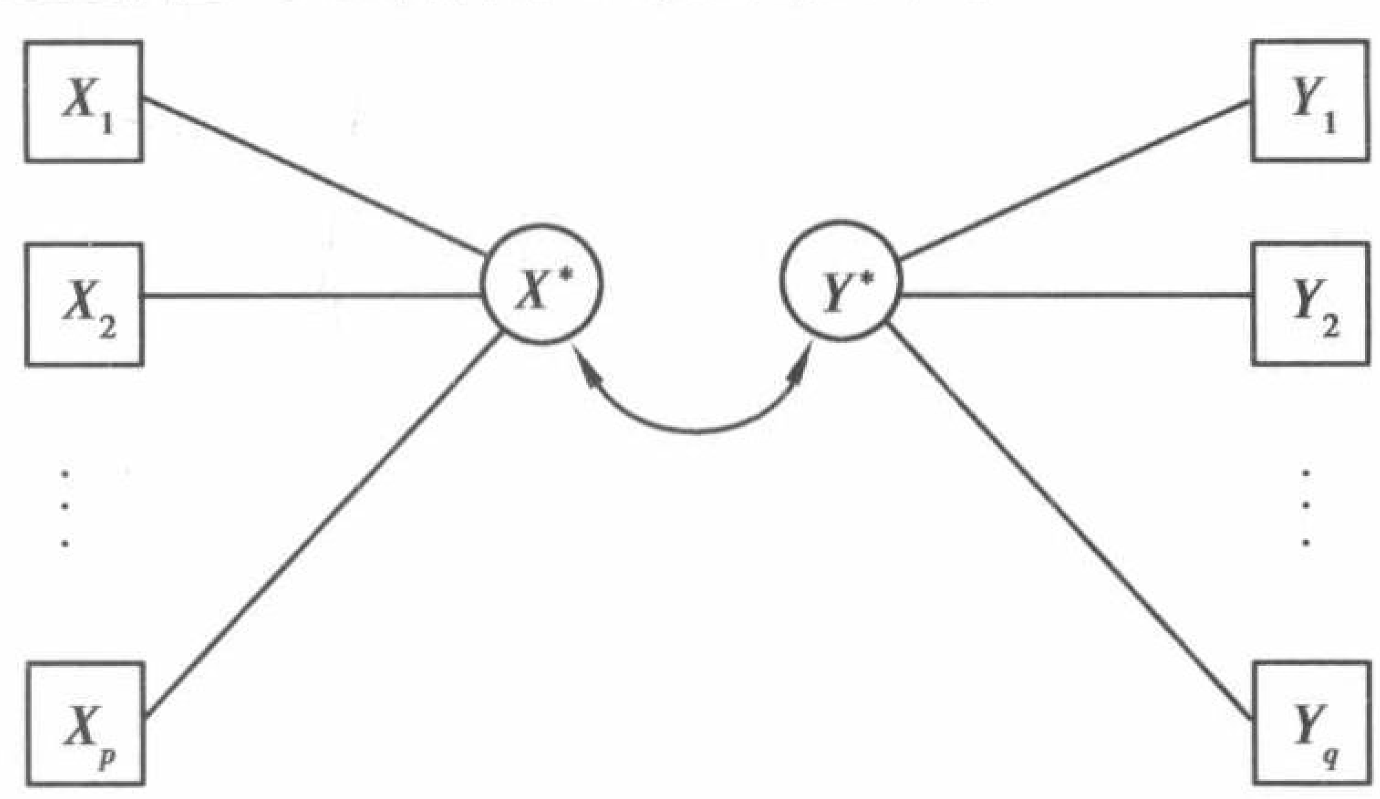



**图1 典型相关**

在拉西特(Russett,1969)研究的问题中,典型变量$X^*$和$Y^*$之间实际上是一种因果关系,因为他认为高经济不平等性的国家将显示高的政治不稳定性。与之相反,我们可以认为政治特性将会对经济特性有影响。但这种观点却未包含在托克维尔或拉西特的观点中。然而,考虑典型相关是一种"相关"这一事实,统计分析并非不对称的。正因为如此,在图中我们并未在从$X^*$到$Y^*$画一个因果的箭头,而是画了双侧曲线箭头,以表明因果关系【Causality】仍然还是开放的。

## 典型相关分析模型

在典型相关分析中我们想看一下一组 $X$ 和一组 $Y$ 之间是否显著关联。为此，我们要找这组 $X$ 的一个线性组合，$X^* = a_1X_1 + a_2X_2 + \cdots + a_pX_p$ 和这组 $Y$ 的一个线性组合，$Y^* = b_1Y_1 + b_2Y_2 + \cdots + b_qY_q$ 之间的组最大相关方式。$X^*$ 和 $Y^*$ 这两个线性组合是不可观察的。在我们的例子中它们预先被命名为经济不平等性和政治不稳定性。有时诸如这样一种事先理论，在研究开始的阶段并不存在，因而名字必须在以后设计中给出。

为了找到这两个线性组合（典型变量 $X^*$ 和 $Y^*$），我们必须以这样的方式计算 $a$ 和 $b$ 的权，即典型相关 $\rho$ 必须尽可能的高。这个典型相关的平方是一组方差（例如政治特性）为另外一组方差（例如经济特性）解释的比例。

恰如主成分分析【Principal Components Analysis，PCA】那样，其中不是一个，而是几个不相关的成分可以被确定，在这里则是各种不同的典型变量对被我们发现。每个对都和前面的对不相关，且每次都要计算，以使典型相关是最大的。典型变量对的数目，最终将与典型相关的数目相同，它等于最小的那组的变量数，即 $\min[p, q]$。

每个典型相关系数及 $a$ 和 $b$ 的权的计算则通过考查典型矩阵的特征值结构来完成。在判别分析【Discriminant Analysis】和主成分分析中，这一典型矩阵分别是 $\mathbf{W}^{-1}\mathbf{B}$ 和 $\mathbf{R}$。在典型相关分析中，这种矩阵是 $\mathbf{R}_{yy}^{-1}\mathbf{R}'_{xy}\mathbf{R}_{xx}^{-1}\mathbf{R}_{xy}$。在这个矩阵记号中，对之内和之间不同类型矩阵作了区分。我们要计算之间相关（$\mathbf{R}'_{xy}$ 和 $\mathbf{R}_{xy}$），但要控制之内相关（$\mathbf{R}_{yy}^{-1}$ 和 $\mathbf{R}_{xx}^{-1}$）。下面我们对此作进一步的解释。

## 三种类型的相关

考查经济和政治特性之间相关的一种合理的方法是分析二者之间单独的相关，例如，基尼系数和领导的（不）稳定性（$X_1, Y_1$），或农民的百分比和暴力程度（$X_2, Y_2$）之间的相关。这些都是 $X$ 和 $Y$ 的变量之间的相关，因此称为之间相关（between correlations）。

而典型相关则比这种考查单独相关的方法更进一步。

首先，在典型相关分析中，我们的目的则在考查两组——一组是 $X$ 变量，而另一组则是 $Y$ 变量——“作为一个整体”，其现实的关联到了什么程度。为了这个目的，我们用一个线性组合取代每组变量。这两个线性组合之间的相关才是我们研究的目标。

第二，在典型相关分析中，我们控制每个组内的关联，这之所以可能，因为在经济特性 $X$ 组内的变量也是成对相关的。这种类型的相关叫作之内相关（within correlations）。它们形成了一个类似多元回归分析中的多重共线性问题。在多元回归分析中，我们计算的回归系数是偏回归系数（即我们控制了 $X$ 组中的每一个变量与其他变量的关联）。在典型相关中我们也要做类似的工作，只是在这里我们不仅要控制 $X$ 组这一方，而且也要控制 $Y$ 组那一方，因为我们也有各种 $Y$ 变量，它们也可能是成对相关的。

因此我们可以区分三种类型的相关：

1.$X$ 变量之间的相关，它的相关矩阵是 $\mathbf{R}_{xx}$；

2.$Y$ 变量之间的相关，它的相关矩阵是 $\mathbf{R}_{yy}$；

3.$X$ 和 $Y$ 变量之间的相关，它的相关矩阵是 $\mathbf{R}_{xy} = \mathbf{R}'_{yx}$。

典型相关分析旨在考查之间相关而非之内相关这一事实，可以通过把它与双主成分分析（PCA 两次）作比较，而予以更为清晰地阐述。PCA 寻找一组变量的内部的相关（也就是内相关）。如果我们做两次 PCA，一次在 $X$ 组，另一次在 $Y$ 组，我们将在每次的 PCA

都得到第一主成分。这两个主成分之间的相关,绝对不同于(第一个)典型相关,因为两个主成分之间的相关测量的是,在多大程度上经济特性组内部的关联可与政治特性组内部的关联相比。与此不同的是,典型相关测量的是,在既控制(也就是假定没有)经济组的之内关联,也控制了政治组的之内关联时,经济和政治特性的相关程度。

## 几何法

为了从几何的观点解释典型相关,我们来建立两个坐标系统:一个以 $X$ 变量为轴,另一个以 $Y$ 变量为轴。典型相关的目的在于找到第一对典型变量,在 $X$ 变量空间中的一个 $X^*$ 和在 $Y$ 变量空间一个中的 $Y^*$,而使用的寻找方法则要使找到的 $X^*$ 和 $Y^*$ 之间的相关最大。然后我们再开始找第二对典型变量,它与第一对是不相关的,但(典型)相关则要尽可能地大。第一对 $X_1^*$ 和 $Y_1^*$,以及第二对 $X_2^*$ 和 $Y_2^*$ 的高典型相关,可以从以典型变量为轴的坐标系统上看出。不同对的典型变量——例如 $X_1^*$ 和 $X_2^*$(及 $Y_1^*$ 和 $Y_2^*$)——是不相关这一事实,并不必定意味着它们在 $x$ 变量空间是正交的。但它的确意味着在 $(X_1^*, X_2^*)$ 空间中的回归线有零斜率。

## 使用这种方法的目的

使用典型相关分析的目的有三个。根据拉西特(Russett,1969)的研究,这些目的可以被描述如下:

1.先寻找第一对线性组合,一个以经济变量组为依据,另一个以政治变量组为依据,寻找的方法是使两个线性组合之间的相关最大。接着再寻找第二对线性组合,也要使相关最大,且和第一对是不相关的。我们连续不断地这样做的次数,以最小组的变量数目为限。我们把这样的线性组合叫作典型变量(canonical variable)。这样的加权叫作典型加权(canonical weight)。这样的最大化的相关叫作典型相关(canonical correlation)。我们试图以典型权数为根据,来对组"之间"的关联作出解释。我们也可以尝试对每一个典型变量的每一对加以命名。

2.利用典型权数,计算典型变量的得分。原变量和典型变量之间的相关,常常给我们提供额外解释的可能性。这些可能性叫作结构相关。

3.对得到的结果进行检查,以断定两组变量之间的关联是否可以推广到国家的总体。这一目标可以通过典型相关的显著性检验来实现。我们既可以对所有的典型相关一起进行检验,也可以对每一单独的典型相关或它们的子组进行检验。因为第一个典型相关是最大的,而随后的相关会变得越来越小,这似乎说明更恰当的做法是,对各个典型相关分别进行检验,进而只对那些最为显著的结果进行解释。

## 因果路数的典型相关分析

1968 年,斯图尔特和洛夫开发了典型相关分析的一种非对称版本。上面介绍的典型相关计算出来的终究都是对称的相关系数。例如,第一个典型相关系数(即第一个特征值 $\lambda_1$)的平方,它表示第一对典型变量之间的相关的平方,可以解释为 $Y$ 组(政治特性)的方差为 $X$ 组(经济特性)解释的比例。然而我们也可以简单地找一条理由反过来说,$\lambda_1$ 也是经济不平等性的方差为政治不稳定性所解释的比例。换言之,这里既没有区分因变量和自变量,也没有区分原因和结果。

为此,斯图尔特和洛夫提出了重叠指数(index of redundancy)。这个指数测量的是以 $X$ 组的知识为根据,$Y$ 组可被重建的程度。这似乎在说,这一系数等同于用一个在整组 $X$ 的函数上的 $Y$ 变量,反复进行多元回归分析得到的所有多重决定系数。

## 显著性检验

为了检验典型相关是否显著,我们将利用威尔克的 $\lambda$(lambda)。这种量度就像学生 $t$,霍特林的 $T$ 和费希尔的 $F$ 一样,将样本之间和之内的离差比率与虚无假设下的这一比率进行比较。威尔克的 $\lambda$ 是这一比率在它最广义的形式中的多变量模拟。所以 $t$, $T$ 和 $F$ 都是 $\lambda$ 的特例。因此这一量度可用于更为复杂的多变量分析方法,如多元判别分析、多变量方差和协方差分析,以及典型相关分析。

在典型相关分析中,威尔克的 $\lambda$ 可以通过计算误差方差的积求得。其中误差的方差则由特征值(即典型相关的平方,因而它就是被解释的方差的比例)减 1 来确定。巴特利特(Bartlett,1947)创建了一种 $V$ 量度,它的分布与卡方分布相同,有自由度 $pq$ 个,其中 $p$ 和 $q$ 分别为 $X$ 组和 $Y$ 组的变量数。

$$V = -\left[n - 1 - \frac{p + q + 1}{2}\right] \ln \lambda$$

对 $pq$ 个自由度而言,在显著水平 $\alpha = 0.05$的虚无假设下,这一值大于临界的卡方值,那么经济不平等性和政治不稳定性之间便是显著关联的。这样,托克维尔和拉西特的假设便在经验上得到了支持。

$V$ 量度也是可加可分的,所以我们可对第二个及后面的经济和政治特性的相关是否显著,单独进行显著性检验。

## 扩展

在实际的多变量研究中,可能会出现这样的情况,即研究的问题由同一单位的两组以上特性组成。例如,在经济 $X$ 和政治 $Y$ 之外,还有一组社会文化特性 $Z$。广义的典型相关这一术语,涵盖了用于研究三组或更多组之间相关的程序。

在前面的讨论中,我们理所当然地认为,所有的变量都是定距和定比层次的。它也同时假定关系都是线性的,而这两个必要条件是可以放宽的。可用于作典型分析的计算机程序 CANALS,也可以作非量度(nonmetric)及非线性的典型分析。计算机软件的这种新发展应该受到我们的重视。对此有兴趣的读者可进一步阅读吉非(Gifi,1980)的著作。

——Jacques Tacq

(沈崇麟译校)

## 参考文献


Bartlett, M. S. (1947). Multivariate analysis. *Journal of the Royal Statistical Society, Series B, 9*, 176-197.

Cooley, W., & Lohnes, P. (1971). *Multivariate data analysis*. New York: John Wiley.

Gifi, A. (1980). *Nonlinear multivariate analysis*. Leiden, The Netherlands: Leiden University Press.

Green, P. (1978). *Analyzing multivariate data*. Hinsdale, IL: Dryden.

Russett, B. (1969). Inequality and instability: The relation of land tenure to politics. In D. Rowney & J. Graham (Eds.), *Quantitative history: Selected readings in the quantitative analysis of historical data* (pp. 356-367). Homewood, IL: Dorsey.

Stewart, D. K., & Love, W. A. (1968). A general canonical correlation index. *Psychological Bulletin, 70*, 160-163.

Van de Geer, J. (1971). *Introduction to multivariate analysis for the social sciences*. San Francisco: Freeman.


# CAPI

参见计算机辅助数据收集法【Computer-Assisted Data Collection】。

## 捕捉-再捕捉法(Capture-Recapture)

捕捉-再捕捉法,也叫作标志-再捕捉法(mark-recapture),是一种估计总体【Population】大小的方法。具体做法如下:捕捉并标记某些总体成员。从总体抽取一个随机样本【Sample】(再捕捉),计数样本中带有标记的成员数。用样本的标记分数估计总体的标记分数。估计的总体大小是总体的标计数除以样本的标记分数。例如,假设总体中标记成员数为1 000,而在一个容量为 500 的随机样本中,标记成员数为 200。那么由此可知,1 000 个总体标记成员大约构成了总体的 200/500=0.4,因而总体的大小便等于1 000/0.4=2 500。法国天文学家、数学家拉普拉斯(Laplace),于 1786 年(Cormack, 2001)第一次使用捕捉-再捕捉法来估计总体的大小。捕捉-再捕捉法有几个关键的假定:

1.样本(再捕捉)是总体的一个简单随机样本。这样每一个总体成员都有相同的入选样本的机会,且一个给定的成员入选样本的机会不能取决于给该成员是否做了标记。这一假定叫作独立性假定,违反这一假定会产生相关偏倚。

2.标记和未标记个体的总体在捕捉和再捕捉这一段时间内,要保持恒定(没有新生、死亡、迁入或迁出)。

3.标记不能含混不清。

以后统计学家又对这种方法作了改进,使它能用于总体有变化、捕捉不等概率【Probability】、时变性概率、“陷阱效应”等场合(Cormack, 2001; Otis, Burnham, White, & Anderson, 1978; Pollock, Nichols, Brownie, & Hines, 1990)。捕捉-再捕捉法有时用于估计动物总体的大小,同时它也一直被应用于估计其他的总体参数,如存活率(Lebreton, Burnham, Clobert, & Anderson, 1992)。捕捉-再捕捉法是建议用来调整美国人口普查存在的少计问题的方法基础(参见普查调整【Census Adjustment】)。

——Philip B. Stark

(沈崇麟译校)

### 参考文献

Cormack, R. (2001). Population size estimation and capture-recapture methods. In N. J. Smelser & P. B. Baltes (Eds.), *International encyclopedia of the social and behavioral sciences* (Vol. 17). Amsterdam: Elsevier.

Lebreton, J.-D., Burnham, K. P., Clobert, J., & Anderson, D. R. (1992). Modeling survival and testing biological hypotheses using marked animals: A unified approach with case studies. *Ecological Monographs*, *62*, 67-118.

Otis, D. L., Burnham, K. P., White, G. C., & Anderson, D. R. (1978). Statistical inference from capture data on closed animal populations. *Wildlife Monographs*, *62*, 1-135.

Pollock, K. H., Nichols, J. D., Brownie, C., & Hines, J. E. (1990). Statistical inference for capture-recapture experiments. *Wildlife Monographs*, *107*, 1-97.

## CAQDAS

计算机辅助定性数据分析软件(Computer-Assisted Qualitative Data Analysis Software, CAQDAS)是一个由菲尔丁和李(Fielding & Lee, 1998)提出的术语,它涉及

广泛的目前能得到的支持定性研究【Qualitative Research】中各种各样分析风格的软件。软件本身并不会“做”分析，它只是辅助。软件需要人们阅读、理解和解释数据。定性数据分析的一个基本部分是有效地处理数据，而这要求我们仔细和综合性地管理大量的文本、视频、代码、备忘录和注释等。一般的文字处理器、数据库程序和专用的文字管理程序一直被有效地用于这些方面。然而在许多定性数据的分析中，其核心工作是编码【Coding】：用命名的思想或概念连接一个或多个不同文件中的一个或多个文本的短文（或视频系列）。第二代CAQDAS引进了编码和操纵的设备，搜索并报告文本、图像或视频中已经编码的内容。这样的编码和检索功能已经成为目前被普遍采用的程序的核心。

## 程序的功能

大多数程序都是用于文本数据的。尽管我们不一定要转录民族志的注释、元叙事等，但这毕竟是它通常所具备的一个长处，因为大多数软件都支持通过显示原文，对提取和/或编码的短文实现快速的语境重构。大多数程序都支持文本的编码，而这样的代码可作为分析的心得，进行分层编排、修改和重新整理。有些软件（如MaxQDA，Qualrus，ATLAS.ti和NVivo）可以用于包括不同的字体、字样和颜色的多信息文本（rich text）。有几种软件支持数字化的音频、图像和视频的分析，有时还可以对数字化资料直接进行编码（HyperRESEARCH，ATLAS.ti，CI-SAID，AnnoTape，the Qualitative Media Analyser）。迄今为止，我们还没有一种软件可以自动地转录音频记录。即使最好的语音识别软件也仍然要依靠高质量的录音，以及对某一个说话者语音的熟悉，因而无法处理多个调查对象的声音。

大多数软件允许写些备忘录【Memos】、编码的定义、田野笔记【Fieldnotes】和其他分析型的文字，作为分析过程的一部分，有些软件（如NVivo，NUD*IST/N6，ATLAS.ti和Qualrus）则能使这些文本链接到项目数据库中的数据和编码。许多程序的一个重要性质是可以搜索在线文本的词句。通常这样的词汇搜索可以把文档中发现的所有与之匹配的文本都标识出来。有些程序支持用编码的文本搜索数据库。其中最为高端的（如NVivo，MaxQDA，WinMAX，ATLAS.ti，NUD*IST/N6，HyperRESEARCH，and Qualrus）使我们能做很复杂的搜索，如涉及布尔代数、邻近的组合词、短语、编码的文本和变量的搜索。

第三代软件又从两个方向扩展了这些功能。有些辅助程序用各种手段考查文本的性质和关系。诸如这样的软件（Qualrus，HyperRESEARCH，Ethno和AQUAD Five）常被称为理论生成器或模型生成器，因为它们含有各种工具，它们帮助研究者发展他们的理论思想和检验假设（Kelle，1995）。某些程序还有建模的功能，它使用图表来链接编码、文本或图像的数据库（如ATLAS.ti，Qualrus和NVivo）。而某些程序则能够实现合作的研究者之间数据和分析的交换。有些程序（如Tatoe和ATLAS.ti）在实现这一目标时是通过共同格式，如XML进行数据和分析的输入和输出。此外，有些程序还能合并不同研究者之间的项目数据库（NVivo，ATLAS.ti和NUD*IST/N6）。

## 软件的挑选

正如韦茨曼和迈尔斯（Weitzman & Miles，1995）所言，各种CAQDAS程序的区别在于设备和操作。哪一种程序更适合，在某种程度上讲取决于各种实际的问题，如资金和掌握的专门知识，不过所需的分析类型

似乎也在必须考虑之列。有些软件程序(NUD* IST /N6)比较适合大型结构性内容的大型数据集和与定量数据库的链接。有些软件允许复杂的分层和多种编码的文本，有些则只能做简单的短文编码。分析风格也是要考虑的重要问题。风格问题涉及主题分析，如扎根理论【Grounded Theory】和各种各样的现象学【Phenomenology】主题都有很好的定性软件支持。然而支持基于个案的资料、叙事分析【Narrative Analysis】、话语分析【Discourse Analysis】和谈话分析【Conversation Analysis】的软件则比较少。

## 长处和不足之处

定性数据似乎都是连篇累牍的，因而很容易产生片面和有偏倚的分析。CAQDAS 的使用对克服分析的片面性是相当重要的，因为它可以去掉许多空乏的分析。它比较容易详尽无遗地分析文本，检查不符案例【Negative Cases】，进而确保文本已被首尾一致和定义精确的方式编码。我们还能保持对分析的审查跟踪，这样可以对已经形成的分析思想进行回想和调整。大多数程序都可以做一些简单的统计分析，这样研究者就可以检查(和给出证据证明)，例子是共同的还是特殊的。一般来讲，CAQDAS 的长处不仅在于它可以进行团队工作、使用大的数据集，或使用音频和视频资料，这些工作都是其他任何方式不能胜任的，但问题还不尽于此，如果没有定性分析软件，诸如这样的事情做起来，将会是令人难以置信的旷日持久。

然而 CAQDAS 的使用仍然还是有争论的。早期的使用者感觉，在使用基于纸质的分析而非计算机辅助的分析时，他们更接近调查对象的话语，或他们的田野笔记。这可能是因为用户比较熟悉传统的路数，对计算机还比较陌生。大多数软件都有优良重构语境的设备，因而对那些使用 CAQDAS 的研究者而言，很少会感到他们与数据之间的距离。还有一种批评意见认为，CAQDAS 会强化某种偏倚，如为扎根理论及编码和回取路数所左右。尽管许多软件的开发都受到扎根理论和其他路数的启发，这使它们在文本编码中占有首要的地位，但现在程序已经变得更为复杂和灵活，且已变得很少专注于任何一种单一分析套路。尽管如此，大多数已出版的研究有关 CAQDAS 使用的文字表明，定性软件的使用至今仍停留在非常简单的范式分析的水平，且很少证明分析的信度和效度。不过随着使用者变得更加有经验，以及有更多的研究者使用这种软件，这种情况有可能发生改变，创造性地使用这种软件的情况必将增多！

——Graham R. Gibbs

(沈崇麟译校)

## 参考文献

Fielding, N. G., & Lee, R. M. (1998). *Computer analysis and qualitative research.* London: Sage.

Kelle, U. (Ed.). (1995). *Computer-aided qualitative data analysis: Theory, methods and practice.* London: Sage.

Weitzman, E. A., & Miles, M. B. (1995). *Computer programs for qualitative data analysis: A software source book.* Thousand Oaks, CA: Sage.

# 延滞效应(Carryover Effects)

在实验治疗实施很久之后，治疗仍继续影响着参与者时，延滞效应便发生了。延滞效应是普遍存在的，因为实际上任何发生在一个机体上的事情都会以某种方式在一段时间内继续影响机体。好比一个池塘表面的涟漪，从扔入其中的石头的入水点向外

散发出去。有些治疗产生的滞后效应数量上相当温和,和/或在持续的时间上相当有限;但另外一些治疗则可能会产生大量和/或持久的延滞效应。

处方药的延伸影响给我们提供了一个最可理解的延滞效应的实例,因为人们十分清楚药物在使用之后的一个相当长的时期内仍继续产生着影响。实际上许多药品生产商的广告都声明他们生产的药品的延伸影响。人们也清楚因不同药物的相互作用而产生的可能的并发症,这使得避免在一种药品仍然存在于他们机体的时候使用另一种药物。

研究者可以采用各种策略,在他们自己的研究中去除或最小化延滞效应的影响。最为明显的策略可能是用独立组设计来消除延滞效应的影响。在这样一种设计中,施于每个参与者的治疗只有一种,因此数据收集完成了,参与者在研究中承担的角色也完成了。因此发生的任何延滞效应,将不会再影响在其他研究条件下收集的数据(尽管治疗可能仍继续影响着参与者)。

在实验中使用重复测量设计【Repeated-Measures Design】时,延滞效应受到了极大的关注,因为参与者受到一系列治疗,在每次治疗之后都要收集数据。这样早期治疗的延滞效应可能继续影响参与者的后续治疗后收集到的反应。当然,在这样的情况下,后来的反应不能完全被解释为那次治疗的缘故,相反它是两次治疗的组合驱动的缘故。

为了最小化延滞效应,从而加大当下正在进行的治疗的预期反应的可能性,一个策略是在两次治疗间留出足够的消逝时间。由于将后一次治疗的时间推迟到足够迟,参与者需要在下一次治疗实施前回到预处理状态。然而,困难在于了解究竟要等多长时间延滞效应才会消退。不仅如此,如果延迟要求参与者几天以后返回,那么(参与者的)损耗便会是一个问题。

还有另外一种基本策略是平衡那些参与者接受治疗的次序,它可以与治疗之间的某些延期组合使用。这样的平衡【Counterbalancing】确保任何延滞效应,以及次序效应【Order Effects】(例如实践或疲劳效应)对所有情况的影响都是相同的,从而消除治疗与位置之间的混淆。

——Hugh J. Foley

(沈崇麟译校)

## 参考文献


Girden, E. R.(1992). *ANOVA: Repeated measures* (Sage University Paper Series on Quantitative Applications in the Social Science, 07-084). Newbury Park, CA: Sage.

Howell, D. C. (2002). *Statistical methods for psychology* (5th ed.). Pacific Grove, CA: Duxbury.

Maxwell, S. E., & Delaney, H. D. (2000). *Designing experiments and analyzing data: A model comparison perspective.* Mahwah, NJ: Lawrence Erlbaum.


# CART

英文 Classification and Regression Trees 第一个字母的缩写 CART,一般都在数据挖掘行业(data-mining industry)使用,它有两个含义:(1)泛指所有的生长分类(无监督学习)法和回归(监督学习)树;(2)指一组特定的生长二叉树算法。该方法由布雷曼等人(Breiman, Friedman, Olshen, & Stone,1984)首先在他们的著作中提出。后者基于一种复杂的交叉验证方法,目的在于调整多分叉树成长算法,如 CHAID(参见本书词条 CHAID)对照的双分叉局限引起的问题。

——Jay Magidson

(沈崇麟译校)

## 参考文献

Breiman, L., Friedman, J., Olshen, R., & Stone, C. (1984). *Classification and regression trees.* Monterey, CA: Wadsworth.

## 笛卡儿坐标(Cartesian Coordinates)

笛卡儿坐标由两条或三条轴(数字线)组成,这些轴彼此成直角。在两维坐标中,$x$和$y$这两个字母常被用来命名两条轴,而在三维坐标中则增加了第三条轴$z$。数据点代表的个案位于坐标中,这样我们便可由此知道数据的分布和由轴表示的变量之间的关系。

——Tim Futing Liao

(沈崇麟译校)

## 个案(Case)

个案是一种研究单位。在总量分析中,例如,在有关美国各州的福利开支的分析中,个案就是州,因而共有50个个案。在一个抽样调查中,被调查者就是个案。在实验【Experiment】研究中,那些被试就是个案。总而言之,样本【Sample】数量($N$)说明了个案的数量,那个$N$中的每一个单位都代表一个个案。

——Michael S. Lewis-Beck

(郑晓娟译　沈崇麟校)

*也可参见个案研究【Case Study】。

## 个案控制研究(Case Control Study)

个案控制研究是用来估计较大且异质的群体的不同因素如何影响特定结果的(发生)风险的一种研究方法。这个方法发端于流行病学研究,用来确认那些与增加各种疾病感染风险相关的因素(Rothman, 1986)。其实,这种研究方法似乎也非常适合社会科学中的问题,因为社会科学研究者经常致力于研究许多国家中各种因素如何影响某些事件,比如战争发生的风险。

在社会科学领域中,过去对于个案控制研究方法的开发利用不是太多,这或许是因为这个研究方法源于流行病学,因此被置于大多数社会科学家的学术训练之外。这种情况确实令人遗憾,因为个案控制研究有几个长处,使得它特别适用于大规模的跨国数据集的研究。

个案控制研究法本身是非常简单的。当我们研究一些诸如战争或革命等稀有事件时,如果研究者感兴趣的是先前的和平时期中诸如这样的事件的发生情况,那么发生数肯定是少之又少。然而,和平的年份的数目极为巨大。一般讲一组既包括和平的年份,也包括了战争和革命开始的年份的“国家年份”数据(由对$X$个国家观察$Y$年所组

成的数据)，将会有数百甚至上千的和平的国家年，而却只有少数几十个事件发生的国家年。个案控制研究法通过将数以千计的和平的“国家年份”作为每一年只有比较少的新信息这种情况来处理、来简化数据分析的过程。因此，与普遍采用的那些将所有的“国家年份”合并到分析中去的大多数统计分析方法不同，个案控制研究法只选取和平年份的一个随机抽样【Random Sampling】，并将它与事件发生年份合在一起。事件发生的年份就是个案，而和平或无事件发生的年份则都是“控制器”。此处，我们还要为每一个个案和控制器设计开发一组自变量。一般个案的自变量都是事件发生的前一年，而控制器的自变量则都是选择的年。然后我们就可以使用传统的统计方法，例如概率对数比或一般最小二乘回归(OLS)来找出那些比控制个案与个案关联更强的变量。

个案控制研究法通常会将“国家年份”数据的规模由数千减少至数百。不仅如此，如果控制个案的选择严格遵循随机原则，那么对自变量影响的估计效果，实际上与采用全部数据集估计的效果几乎相同(King & Zeng, 2001)。

使用大规模跨国数据来研究战争、革命及其他稀有事件的研究者，都面临许多对特定变量影响进行有效推论的障碍。其中两个障碍是操作性困难，另外两个则涉及更加实质性的问题。个案控制研究法在上述问题的解决中都很有帮助。

第一个操作性困难在于大规模的跨国数据集中——经常只有某些种类的数据是可获得的，如人口和经济数据。某些数据(如政体类型)，尽管对于许多国家许多年份都可获得，还有另外一些类型的数据(如政治领导层的特征或某些特定的领导人)需要费时费力地手工编码才能保证数据的完整。还有一些种类的政治单位，如城市、区域或省份，相关资料可能更加有限。对于数千个个案的数据进行收集和编码的代价是难以承受的，而只进行数百个个案的数据整理则更加可行。因此，使用个案控制研究法使我们有机会去分析更多不同的令人感兴趣的变量，而不是仅局限于能够覆盖所有国家、所有年份完整数据的少数变量。

第二个操作困难也与资料的获得有关。即使一些数据声称覆盖了所有国家和所有年份，事实上数据总会有大量的缺失，或者数据是以 5 年或 10 年为间隔提供的。进行常规分析需要将所有国家和所有年份的数据值补齐，因此需要依赖大量使用虚拟估算和数据插入技术。但是，研究者感兴趣的通常正好是那些主要的破坏性事件发生前的那些年份，而这些“国家年份”的数据往往最有可能是不完整的，因而常规的插入或估算很有可能是无效的。其结果是，对尼日利亚人口的了解可能在 20%的精度以内，而对加拿大人口的了解可以到小于 1%的精度。然而如果想检验不同的人口变量对种族冲突的影响，前面的个案却是特别重要的。通过使用个案控制法来减少“国家年份”的数量，研究者不必将每个缺失数据的值都补足，或者对一段长期数据依赖插入和估算技术。研究者可以集中处理从控制个案样本中抽取的特定的“国家年份”，并设法提高它们在特定观察上的精度。

个案控制法所拥有的优势还体现在更为实质的两个方面。第一，当我们在常规的回归分析中使用全部“国家年份”的数据时，大部分国家会拥有长时间连续的和平年份。这会带来序列相关【Serial Correlation】问题。尽管计量经济学方法能够在单个国家的研究中消除时间因素的影响，但目前还没有有效方法处理多个不同国家的研究中的此类问题。由于个案控制法只使用和平“国家年份”中的一段样本数据，在这样的分析中就不会存在长时间连续的和平年份。因此避免了序列相关问题。

第二,当我们使用"国家年份"的全部数据时,经常会带来比较"苹果"和"橘子"的问题。这个说法的意思是,大多数和平的"国家年份"会来自富裕、民主的国家,而大多数战争或暴力发生的"国家年份"则来自贫穷的、非民主的国家。原因很显然,可能因为富裕和民主会导致和平。但也可能是由于世界上那些处于和平状态却贫穷、非民主的国家被我们忽略了,所以此类国家代表性不足。又或者由于那些富裕、民主的国家有一些其他特征导致了他们的和平(比如这类国家聚集在欧洲和北美,远离贫困地区)。在任何一种情况下,使用个案控制法,都能够通过将控制个案和个案进行匹配来解决这些问题。

例如,研究者可以通过对随机抽取的控制个案进行分层的方法,使控制个案在时间、区域、殖民历史或其他因素方面与那些个案进行匹配。这会增加数据中要分析的代表性不足的群体出现的频率,并防止结果由在某些变量上的国家的随机分布所致。对于那些无法通过政策加以改变,但却可能转变一些国家的基线风险,如区域、殖民史或时期的变量,这样的匹配尤为合适。

无论控制个案是简单随机选择的,还是以某种标准匹配的,其数目都既可以与研究个案相同,也可以是研究个案的几倍,即为每一个研究个案选择多个控制个案。重要的是要有足够进行显著性统计分析的观察,而不是拘泥于研究个案与控制个案的某个比率数。因此,如果我们有 500 个个案,研究个案与控制个案 1∶1 的比率对大部分研究目的而言都可以构成足够大的样本 $N$;但如果只有 100 个个案,研究者可能希望为每一个研究个案选择 3 或 4 个控制个案以便能形成一个大的 $N$。匹配个案控制研究法可能需要有更多的控制个案,以使每一层都能有足够的研究个案和控制个案数。

流行病学的研究文章(如 Rothman, 1986)为我们提供了适合用个案控制样本进行估计的方法。社会科学家应该更多地利用这些技术。如果这些方法能够帮助医学分析者确定在大规模个体的总体中,究竟哪些因素引发癌症或其他疾病,它们也能够帮助社会科学家确定,在不同社会的大规模总体中,究竟是哪些因素导致了战争或其他形式的暴力事件。

——Jack A. Goldstone

(郑晓娟译　沈崇麟校)

## 参考文献

King, G., & Zeng, L. (2001). Logistic regression in rare events data. *Political Analysis*, *9*, 137-163.

Rothman, K. J. (1986). *Modern epidemiology*. Boston: Little, Brown.

# 个案研究(Case Study)

个案研究这个术语的使用并不是十分规范。从字面意思理解,它的使用容易引起误解,因为感觉上所有研究都是对个案的调查分析。但是,我们还是要明确这个术语的核心含义,个案研究是对少数个案,甚至可能只是单个个案的相当深入的研究。当然,个案研究还有一些其他的特征也常被提及。

个案研究法经常与社会研究设计中另外两种非常有影响力的方法——调查【Survey】和实验【Experiment】放在一起比较。与社会调查方法相比较的那些维度已提及的有:调查的个案数和研究者收集的每

个个案的信息详尽程度。在其他条件相同的情况下,研究的个案数越少,对每个个案所收集的信息就能够越详尽。社会调查研究数量众多的个案,但是对每个个案通常只收集相对较少的数据,即集中在个案的某些特定方面(社会调查中的个案大部分情况下都指的是个体被访者,当然也有例外)。个案研究与此不同,研究者会收集一个或较少个案的大量信息,关注个案多方面的特征。个案研究中的个案可以是个人(例如在生活史【Life History】研究中)、事件、机构,甚至整个国家或某个地区。我们遇到的难题是,当要研究的每个个案较大时,我们通常要使用社会调查技术来收集个案中的数据。

个案研究与实验研究也有显著区别。尽管实验研究与社会调查相比,也是对较少数量的个案进行研究,它与个案研究的区别在于对变量的直接控制。在实验中,研究者创造了所研究的个案(或多个个案),而在个案研究中研究者是在自然发生的社会现象中发现个案。这也是一个维度,而非一个二元对立,准实验【Quasi-Experiment】和田野实验法【Field Experiment】则处于这个维度的中间位置。

个案研究这个术语既表明了收集资料的种类,也表明了资料分析的方法。通常情况下,个案研究意味着收集非结构化的资料,并对资料进行定性分析。此外,它还与一个有关研究目的的更为基本的议题相关联。有时研究者认为,个案研究是为了发现个案的独特性,而不是以个案为基础得出更广泛的经验性或理论性结论。这种说法再次强调了:个案研究并非绝对地把得出普遍性结论的意图排除在外,但它意味着研究者是通过在特定个案中的发现用推理而得出那些结论,而不是通过选出的个案来检验假设。有鉴于此,经常有人认为,个案研究应该采用一种归纳【Inductive】的取向,但是并非所有的个案研究者都接受这种观点。

另外一个与个案研究相关的问题是客观性【Objectivity】,至少在这个术语的某种意义上情况如此。研究者在对每个个案进行解释时,究竟应该采取外部视角,或者说是与个案中的被研究对象截然不同的研究观点?还是应该单纯地"以他们的方式"呈现每个个案的特征?当研究个案是人时,这个差异体现得最为明显,因为研究目的可能变成了让他们"发出声音",而不是把他们作为研究的参与者甚至是被研究对象(respondents)。然而,尽管这个区分看起来十分清晰,实践中却往往复杂得多。同一个个案中不同的参与者可能有不同的观点,甚至同一个人在不同情况下也会表现出不同的观点。另外,确定参与者所呈现的东西是否已经"抓住"了他们的观点,而不是他们"外在"的表象,这个问题也相当复杂。

一些评论者认为,一个个案总是某种重要事情的一个"个案",因此对某种一般范畴的兴趣从研究一开始便已隐含其中,尽管那个范畴可以随着研究阶段的进展而有所变化。正因为如此,有些个案研究的调查路径,会利用比较法【Comparative Method】来定义和提炼一些理论范畴。这方面的例子包括扎根理论【Grounded Theory】,使用约翰·斯图亚特·密尔的一致与分歧法的历史研究路径(见 Lieberson in Gomm, Hammersley, & Foster, 2000)以及分析式归纳【Analytic Induction】等。

另外一个关于个案研究的分歧是:究竟应该将个案研究看作一种既有一些长处,也有一些短处,使用的场合和时机要视调查的问题而定的方法;还是一种其取舍只视哲学或政治的原因而定的范式路数(paradigmatic approach)。即使在个案研究被简单地看作一种方法的时候,个案研究采用的特定形式也可能存在很大差异:

- 研究的个案数量;
- 是否有比较,如果有,它起的作用;

• 个案研究的详细程度；

• 个案处理的范围大小；

• 研究者将什么作为个案的背景，他们如何确认背景，以及他们试图记录它多少；

• 个案研究者在描述、解释和/或提出理论或进行评估和/或提出方案的方面对自己的限制程度。

这些方面的差异很大程度上取决于个案研究要达到的目的。如果个案研究用来验证或证明一个理论观点，研究者会将个案看作某种类型的一个例证，用一种特定的理论框架（隐含的或外显的）来描述该个案。如果个案研究是为了发掘或发展新的理论观点，它就可能在处理研究的过程中更加细微深入，并具有更强的开放性。类似的，个案研究的目的是对特定情况下所发生事件的描述还是解释，也会带来研究的差异。如果研究者感兴趣的是调查中发现的问题，那么个案研究的讨论就会转到诊断问题、找出原因，或者还有开出治疗方案等方面。

研究目的的差异还会体现在对调查个案的选择方面。有时研究者会重点关注那些被认为是某个领域的典型，或是某个群体代表的个案。另外一些时候，研究者会寻找异常或极端个案，因为这些个案最有可能激发新的理论观点，和/或提供对某个假设的检验。

然而，有许多评论者认为，个案研究不仅仅是一种方法：对于社会生活能够和应该如何被研究这一问题，它包含了与那些潜藏在其他研究路数下的假定——有相当大差异的一些假定（Hamilton，1980）。有时，我们称实证主义【Positivism】为一方，以自然主义【Naturalism】、诠释主义【Interpretivism】或者建构论【Constructionism】为另一方。有一种极端观点认为，个案研究对社会现实的揭示与小说家、短篇故事作者，甚至诗人的方式非常类似。对个案研究持有这种看法的人会在对比个案研究和其他研究方法时，以根本错误的方式来评判各种方法的优劣势。

由对于个案研究的目的和本质的理解差异引发了一系列的方法论问题，并引起了广泛的讨论。

1.概推性。有些个案研究致力于为得出一些普遍类型的现象或关于更广泛个案群体的结论奠定基础。这就带来一个问题，这种想法是否可能。有些学者认为个案研究所进行的推论或普遍化是“逻辑上”“理论上”和“分析性”的，这与统计分析中的推论本质不同（见 Mitchell in Gomm et al.，2000；Yin，1994）。另外一些学者则认为，个案研究在某些方面可以用来进行与社会调查研究相同意义的普遍化（见 Schofield in Gomm et al.，2000）。还有些学者认为个案研究不是必须对结论进行普遍化，重要的是读者如何利用个案研究的结论——学者将这一过程称为“自然而然的普遍化”（见 Stake in Gomm er al.，2000）。另一种说法是根据“适合程度”促使研究结论从一个情景到另一个情景的“转化”（见 Lincoln & Guba in Gomm et al.，2000）。但是，关于后面这种研究方法，还存在普遍化的有效性问题。

2.因果分析还是叙事分析。个案研究学者有时声称，他们通过检验一两个个案，就有可能确定一些抽样调查研究都无法发现的因果过程。这是因为个案研究不仅局限在某一点上，而且对个案进行了深入和长时间的研究。还有一个常被提及的观点是，与实验研究法处理的是人工制造情景中的因果关系不同，个案研究发现的是“真实世界”中的因果过程。这种观点还有一些表述方式则强调，结果总是可以通过多种路径达成，因而在特定个案中事件发生的叙述式原因，对于我们理解结果的发生是至关重要的（见 Becker in Gomm et al.，2000）。这一点与历史学家的工作并行不悖。但是，如果研究者只研究单个或很少数量的个案，那么如何将事件中的偶然原因，与必然的因果关系

区分开来仍然是一个问题，同时还会有理论【Theory】在因果/叙述分析中究竟起了什么作用的问题（见 Gomm et al., 2000, Chap. 12）。

前面已经提到过，有些个案研究学者声称可以通过对照分析来确认因果关系，例如，通过密尔的一致和分歧法或者通过分析式归纳法。有时对照法被认为与统计分析法有相似之处，但更多时候“统计分析”和“个案研究分析”在“逻辑”上仍然具有显著差异。不过，对于在逻辑上二者是否具有差异也同样有人提出了疑问（Robinson in Gomm et al., 2000），同时也有人质疑密尔的准则或分析式归纳法是否能够作为通过个案研究产生理论的合适方法（Lieberson in Gomm et al., 2000）。

3.理论的本质。尽管许多个案研究学者都强调了理论的作用，但是在所需理论视角的特征这一问题上观点则不尽相同。对某些人来说，理论必须能使个案成为一个有边界的系统。这里强调的是，我们必须把个案当作只能从整体来理解的独特结构（见 Smith, 1978）。对另一些学者来说，理论的作用更多在于查找和解释，个案在其所处的更为广泛的社会背景中，其内部正在发生什么（Burawoy, 1998）。他们认为，如果没有理论的作用，个案内部过程非常有可能被误解。的确，我们经常看到这样的说法，对个案的分析总是预设某种更广泛的社会背景，这样研究的问题便不是与宏观理论是否有涉，而是它究竟清晰到什么程度，以及它是否言之成理。

4.真实性与权威性。有时，我们提倡个案研究方法的原因在于，它能够捕捉到个人、情景、群体等的独特性质。在这里，我们关心的问题并不是将个案作为一个范畴的代表，或者将个案推广至总体。研究目的在于“以它自己的话语”真实地呈现个案。在这一观点的某些版本中，它被看作发现文学和艺术提供的那种符号性真理的基础（Simons, 1996）。但是，对于这个观点的具体含义仍然是有疑问的，因为不同的美学理论，其取向毕竟是大相径庭的。

对真实性立场的坚守还基于个案研究学者表现出的对任何权威判断的反对，和/或他们所持有的立场，认为个案研究能够扩大那些在大家看来，其经历和观点被忽视的人群的独特声音。但是关于这个立场也存在很多质疑，质疑的人不仅包括那些坚守科学研究路径，或那些强调宏观理论作用的学者，还包括一些建构主义者和后现代主义【Postmodernism】。建构主义者和后现代主义者否认在调查研究之外的真实现象的存在，质疑研究者为他人发声（甚至充当中间人）的合法性，和/或者质疑那种个案研究描述中得到的人们拥有整齐划一观点的理念。通过这些方式，建构主义者和后现代主义者动摇了真实性的概念。

概括而言，尽管近年来个案研究法越来越流行，并且具有相当大的价值，但对这个术语的使用，仍然涉及很多不同的研究路径，与此同时，它还引起了一些基本的方法论问题。

——Martyn Hammersley

（郑晓娟译　沈崇麟校）

## 参考文献

Burawoy, M. (1998). The extended case method. *Sociological Theory*, *16*(1), 4-33.

Gomm, R., Hammersley, M., & Foster, P. (Eds.). (2000). *Case study method*. London: Sage.

Hamilton, D. (1980). Some contrasting assumptions about case study research and survey analysis. In H. Simons (Ed.), *Towards a science of the singular: Essays about case study in educational research and evaluation* (pp. 78-92). Norwich, UK: Centre for Applied Research in Education University of East Anglia.

Simons, H. (1996). The paradox of case study. *Cambridge Journal of Education*, *26*(2), 225-240.

Smith, L. M. (1978). An evolving logic of participant observation, educational ethnography and other case studies. *Review of Research in Education*, *6*, 316-377.

Yin, R. (1994). *Case study research* (2nd ed.). Thousand Oaks, CA: Sage.

## CASI

参见计算机辅助数据收集【Computer-Assisted Data Collection】。

## 突变理论(Catastrophe Theory)

突变理论指的是在某些非线性动力学数学模型中,那些经历非线性动力学【Nonlinear Dynamics】的性状类型,即一个参数(称作控制参数)的值发生微小变化,就导致一个变量的值发生突然或剧烈的大量级的变化。在这个意义上,突变理论也能为那些大体上与"一根稻草压垮骆驼背"类似的情境现象建模,尽管突变理论的应用可能是非常普遍的。当代对突变理论的理解发端于汤姆(Thom,1975)的研究工作。

所有关于突变理论的早期研究几乎都在微分方程数学模型的设定中使用了多项式函数。在某种程度上说,这也是汤姆(Thom,1975)发现的通用性的一个重要结果。因为所有足够光滑的函数都可以用泰勒级数近似(Taylor series approximation)来加以扩展(得到一个初始模型的多项式表达),直接对这个多项式表达进行分析也是可能的(Saunders, 1980, p. 20)。不过,科学家们也可以避免使用某一个汤姆经典多项式,而使用他们自己的有充分理论根据的原始设定,只要他们清楚原始设定存在突变的可能(Brown, 1995a)。

突变理论的基础是一种有关分叉(bifurcation)的观念。分叉是一个动态系统进化过程中发生的事件,在这个过程中,系统的特征行为发生改变。当系统中的一个吸引子(attractor)随着一个参数(称作控制参数,因为它的值控制着突变现象)的值改变而改变时,分叉就发生了。突变是一种分叉(与微小的分叉,或爆炸性的分叉比较而言)。动态系统的特征行为是由其行为轨迹决定的,而行为轨迹则是系统中变量随时间变化的值。当轨迹与一个分叉相交时,它们一般会采取一种与受到分叉影响之前的类型完全不同的行为。因此,如果一个轨迹即将紧紧"拥抱"吸引子或系统中的平衡点并与分叉相交,轨迹便可能会突然放弃之前的吸引子,而"飞"到与另一个吸引子邻近的区域。突变的基本特征就是某个吸引子及其吸引盆(basin)突然消失,同时另一个占据主导地位的吸引子出现。因为多维表面也可能有吸引力(连同这些表面上的吸引点一起),因此这些动力系统中的重力中心常被称为超曲面、极限集或简单地叫作吸引子。

下面这个模型是齐曼(Zeeman,1972)提出的,它说明了一个尖端突变,并被用来为一个跳动的心脏中的肌肉纤维长度(变量 $x$)的变化建模。控制参数 $A$(在这个例子中指的是最终下达指令指示心脏何时跳动的电化学活性)的值可能会持续变化,它被用来移动轨迹,穿越具有突变可能的均衡的超曲面。参数 $q$ 标明了系统中的总

张力,而 $f$ 则是尺度参数。系统里的两个不同的方程分别是:$dx/dt=-f(x^3-qx+A)$ 和 $dA/dt=x-x_1$。在这里 $x_1$ 代表的是心脏收缩(心脏收缩平衡)时的肌肉纤维长度。设导数 $dx/dt=0$,根据系统其他参数的值,我们将发现 1~3 个 $x$ 值。在一个给定的控制参数 $A$ 的值有三个 $x$ 平衡点时,其中有一个平衡点是不稳定的,不会吸引任何轨迹。另外两个平衡点会为了吸引附近的轨迹而相互竞争,当有一个轨迹经过系统中的分叉点时,它会放弃其中一个平衡点,迅速改变位置,进入另一个平衡点的邻近领域(也即吸引盆)。这种轨迹的迅速改变就是突变。

布朗(Brown,1995b, chaps. 3 and 5)开发和探讨了非线性微分方程动态突变设定(specifications)的两个社会科学的实例。一个例子与美国 1980 年总统选举有关,讨论的是对候选人的倾向、对某一政党的感觉以及个人的政治背景或环境质量三个因素之间的相互作用。另一个例子讨论的是魏玛共和国选民中的党派分裂、选举的去机构化,以及对纳粹党的支持三个因素之间的相互作用。两个例子都经过了充分的估算,第一个例子使用了个人和加总数据,而第二个例子只用了加总数据。

——Courtney Brown

(杨可译　沈崇麟校)

## 参考文献


Brown, C. (1995a). *Chaos and catastrophe theories.* Thousand Oaks, CA: Sage.

Brown, C. (1995b). *Serpents in the sand: Essays on the nonlinear nature of politics and human destiny.* Ann Arbor: University of Michigan Press.

Saunders, P. T. (1980). *An introduction to catastrophe theory.* New York: Cambridge University Press.

Thom, R. (1975). *Structural stability and morphogenesis.* Reading, MA: Benjamin.

Zeeman, E. C. (1972). Differential equations for the heartbeat and nerve impulse. In C. H. Waddington (Ed.), *Towards a theoretical biology* (Vol. 4, pp. 8-67). Edinburgh, UK: Edinburgh University Press.


## 分类(Categorical)

在分类这一术语中包含诸如定类变量【Nominal Variable】和定序变量这样一些术语。大体上讲,这一术语的名称源于一个事实,即一个分类变量无非就是把观察分类归入离散的组群,如把样本的宗教信仰分类归组。

——Alan Bryman

(沈崇麟译校)

*也可参见属性【Attribute】、离散【Discrete】、定类变量【Nominal Variable】。

## 分类数据分析(Categorical Data Analysis)

社会科学的数据在数量化之后可以进行各种各样的统计分析,其中使用最为普遍的是回归分析。然而,它要求因变量是连续的,此外还必须满足回归要求的其他各种假定,这使线性回归难以成为总是能让我们满意的分析工具,因为很多社会科学的数据都是分类的。

社会科学数据可以用两种最为常用的方式进行分类。第一种,在变量以定类或离散的分组记录时,它就是分类的。在政治科学里,一个美国的政治候选人可以是民主党、共和党或独立候选人。在社会学里,人

们的职业是作为一种离散的结果来研究的。在人口学里,避孕工具的选择,如避孕药或避孕套是分类的。教育问题的研究者可能研究中学高年级学生的离散的高等教育目标:大学、社区学院或职业学校。

此外,一个变量可以采用定序尺度,如李克特量表【Likert Scale】或其他与之类似的尺度。诸如这样的尺度在心理学中使用得尤其广泛。在社会科学则一般用它来测量人的特质、态度和观点,通常它有五个定序类别,从最重要到最不重要或从完全同意到完全不同意,二者之间则是中性的中间类。一个定序尺度有两个主要的性质:在各个类别之间存在着一个自然的等级,位于较低位置的类别,与其相邻类别间的距离并不一定在尺度上是等分布的。不是量度个性或态度的变量也可以用定序尺度来表示。例如人的教育程度可以只用三个类别,"小学""中学"和"高中及以上"来测量。这些类别服从一种自然的等级,但是"小学"和"中学"之间的距离,不一定和"中学"和"高中及以上"之间的距离相等。显然,一个定序尺度至少要有三个等级类别。尽管不存在类别总数的上限,但实际上研究者几乎不用七个以上类别的定序尺度。

上面分类数据的例子表明它们在社会科学研究中使用的广泛性。分类数据分析实际上是分类应答(因)变量的分析。

## 历史发展

卡尔·皮尔逊和乌迪·耶尔二人在20世纪之交有关分类变量之间的关联的工作,为离散应答变量模型在以后的发展铺平了道路。皮尔逊的贡献通过与他同名的卡方统计量为众人所周知。而耶尔则强烈倡议在分析关联中使用优势比【Odds Ratio】。尽管知名的统计学家如费希尔和威廉·科克伦在这方面贡献至伟,但分类数据分析能有今天这样的发展,则是在1960年代之后的事情了。

第二次世界大战之后的数十年,我们看到人们对社会问题的解释兴趣不断高涨,因而社会对高技能的社会科学研究者的需求也迅速增长。北美的大学在扩大招生,以容纳战后生育高峰时出生那一代人入学,同时大学内的社会科学教学队伍和研究实体也在迅速成长。分类尺度也应运而生,被广泛用来测量态度、社会阶级及其他许多属性和概念。而与此同时,定量数据也可供教学和研究机构共享。社会科学的方法学也日益变得复杂高深,以满足处理和分析不断增长的可资利用的分类数据的需要。许多开发了离散应答(因)变量回归类模型的重要的统计学家,都来自社会科学附属机构,或是与之有联系的学者、大学教师就不足为怪了。诸如这样久负盛名的统计学家有利奥·戈德曼、谢尔比·哈伯曼、弗雷德里克、莫斯特勒、史蒂芬·芬伯格、克里弗德和克洛格等。这些方法学家专注于对数线性模型的开发使用,而他们在生物科学中的同行,则专注于分类数据的回归类模型的研究。正是这些统计学家,和他们在生物界的同行一起将分类数据分析提高到了一个新的水平。

在20世纪后期的20年,人们见证了学界对于研究定序变量,特别是那些它们的值被假定是潜在或不可直接测量的定序变量之间的关联的巨大关注。在同一时期,人们也目睹了拉扎斯菲尔德于1950年代进行的潜变量模型兴趣在学界的复兴(如潜类分析),并把它们扩展到回归类设置。在这两个领域中的领军人物是利奥·戈德曼、克里弗德和雅克·哈赫纳尔斯,他们是北美和欧洲社会科学方法学的杰出代表。

分类数据分析的另一个传统源自计量经济学。在计量经济学中,研究者使用一种(在观察值中,或在分布中)有这样或那样限制的因变量的回归类模型。马达拉对这种方法的普及居功至伟。在因变量有两个值(二分)、有限个等级数(定序)或无序的(定类)的类别,且是可以观察的时候,Logit和Probit模型便是我们可以使用的恰当模

型。在过去的 25 年,计算机软件用于做二分、定序和多类 Logit 或 Probit 模型的计算机软件有了迅速的发展。在统计学中,许多分类数据模型,如 Logit、Probit、泊松回归,以及线性回归可以作为广义线性模型【Generalized Linear Model】族被更为方便地表示和研究。

## 分类数据分析的类型

有若干种对分类数据分析法进行分类的方法。根据方法的参数性质,我们可将分类数据分析方法分为两类。

1.*非参数法*。这些方法要做最低限度的假定,且对假设检验十分有用。这方面的例子有皮尔逊卡方检验【Chi-Square Test】,费希尔的精确检验(对小的期望频数)和 Mantel-Haenszel 检验(对线性关联的定序分类)。

2.*参数法*。这种基于模型的路数假定采用的抽样方法是随机(可能是分层)抽样,可用于估计的目的,也易于拟合许多特定的模型(例如对称、准对称)。它也可以进行标准误【Standard Error】、协方差【Covariance】和模型参数的置信区间【Confidence Interval】,以及预测的答案的概率估计。

第二种类型在社会科学中的应用最为普遍,这一点有分类数据的参数模型的广泛应用为证。社会研究中的变量在统计模型中既可以是因变量,也可以是自变量。它们也可分别叫作应变量和解释变量。由变量的类型(分类的或非分类的)和所处位置(自变量【Dependent Variable】或因变量【Independent Variable】)产生了一些分类数据分析模型,给我们提供了另一种给分类数据分析法分类的途径。为了简便起见,我们将所有变量视为分类或连续的(即用定距或定比尺度的)。在某些模型中,所有的变量都可以看作因变量,只有它们之间的关联才是研究者感兴趣的。对于那些区分因变量和自变量的模型,我们将注意力限制在那些有单个因变量的模型。这就是说,在这样的模型中,每个模型不论自变量的数目有多少,它们可以都是连续的、分类的或二者的混合,而因变量却只有一个。

关联模型不区分因变量和解释变量。这就是说,这些调查中的变量是彼此相依的。在成对的两个变量都是连续的时候,我们就使用相关法分析,在它们都是分类的时候,我们就使用列联表或对数线性分析。这种情况可以推广到多变量,这就像偏相关分析和多向列联表那样。

在成对的两个变量类型不同时(例如一个是分类的,另一个是连续的),由于要区分自变量和因变量,因此要对因果次序作假定。为了分析连续的因变量在解释变量类别中的分布,我们就要用方差分析或线性回归,这些方法可以推广到包含额外的、连续的解释变量这样的情况。那时,我们就要使用协方差分析和线性回归。Logit 和 Probit 模型的灵活性为它们适用于表 1 中分类因变量的所有三格,不论解释变量的类型是连续、分类还是混合的所证明。常常也会有这样的情况,同一组数据可以用不同的方法进行分析,但究竟用什么方法则取决于研究者想要作的假定和研究的目的。

**表 1　变量类型和统计模型对应表**

| | | 因变量 | |
|---|---|---|---|
| | | 连续 | 分类 |
| 因变量 | 连续 | 相关 | 列联表分析,对数线性回归 |
| | 分类 | | |
| 自变量 | 连续 | 线性回归 | Logit 和 Probit 模型 |
| | 分类 | 连续 | Logit 和 Probit 模型 |
| | 混合 | 协方差分析,线性回归 | Logit 和 Probit 模型 |

因此在表 1 中,标题为分类因变量列中的格都是由分类数据分析法组成的。我们

对这些方法再作进一步的考察，观察这些应变量是否有与纯粹的定类变量不同的有序的类别。有些未在表1中列出的方法，例如泊松回归，这种可给服从泊松分布的整数因变量建模，它也是一种有限因变量模型(limited-dependent variable model)。

依据变量是离散的还是定序的，分类数据模型还可以作进一步的分类。例如在因变量是二值的时候，Logit(或Probit)模型是二分的；在因变量是有序分类的时候，Logit(或Probit)模型是定序的(或有序的)；在因变量有两个以上的离散类别时，Logit(或Probit)模型是多类的。用于分析两个定序变量之间关联的对数线性模型，可以扩展到线性对线性的关联；而对数乘积模型则包含了这二者(更详细的阐述，请参见对数线性模型【Log-Linear Model】和关联模型【Association Model】)。迄今为止，我们只是考查了观察变量的分类数据分析。社会科学家也将分类数据分析用于潜变量，用它们来表示那些不可直接测量的概念。因而在这里，我们还有另一种对分类数据进行分类的方法，即依据模型中的潜变量观察变量，或同时依据二者是否是分类的来进行分类。

分类数据分析尤其关心表2中的第二行，在这一行中的观察变量是分类的。用类别观察变量估计类别潜变量的潜类分析【Latent Class Analysis】，一直受到社会科学界的极大关注。

**表2　与变量类型对应的潜变量模型**

| | 潜连续 | 潜分类 |
|---|---|---|
| 观察连续 | 因子分析模型；SEM | 潜剖面(latent profile)模型 |
| 观察分类 | 潜质(latent trait)模型；SEM | 潜类(latent class)模型 |

若干大型计算机程序使潜类模型为更多社会科学家所接受和使用。这些软件包括潜类分析；最近在明尼苏达大学的伊莱亚森分类数据分析系统安装的MLLSA程序；哈佰曼的NEWTON和DNEWTON；哈赫纳尔斯的LCAG和弗蒙特的LEM等。

## 模型拟合和检验统计量

除上述分类法之外，还有一种看待和对分类数据分析进行分类的方法，这种方法关注模型的拟合是怎样定义的。当今使用分类数据的社会科学研究许多都源自三个虽然一般但彼此关联的传统：一种是将期望频数【Expected Frequencies】与观察频数进行拟合。另一种是将期望的因变量的值与观察的因变量的值进行拟合，最后一种是将期望的关联量度，通过相关与观察的对应量度进行拟合。这第三种传统描述分类数据是如何在因子分析和结构方程建模法中处理的。在本书的其他条目中对它有专门的介绍，在这里我们将集中介绍前两种传统。

第一种传统拟合一种特定的统计模型与一个频数表。令 $f_i$ 为在表中格子 $i$ 中的观察频数，而 $F_i$ 则是在格子 $i$ 中的期望频数。而我们模型拟合的目的是精简观察和期望频数之间差别的某个形式。因此，我们求得皮尔逊卡方统计量和似然比统计量【Likelihood Ratio Statistic】$L^2$(或 $G^2$)：

$$\chi^2=\sum_i\frac{(f_i-F_i)^2}{F_i}\text{ 和 }L^2=2\sum_i f_i\log\left(\frac{f_i}{F_i}\right)$$

和号涵盖全部的格子 $i$，除了那些有结构零的格子 $i$。这些统计量本身可以在频数表的非参数法，或像对数线性模型那样的参数法中单独使用。对数线性法一般将 $\log(F_i)$ 表示为一张表中的行、列和层等与它们的交互项效应的线性组合。上面给出的两个检验统计量，实际上是一个叫作验力散发统计量族(power divergence statistics)的特例。这一统计量族还包含第三个成员，Cressie-Read统计量【Cressie-Read Statistic】，它代表

$\chi^2$和$L^2$之间的一个折中。

第二个传统是广义线性模型,它试图将因变量$y$的期望值$\mu$,通过建模成为解释变量和它们的参数的一个线性组合的函数。我们期望数据服从指数分布,而连接$g(\mu)$与自变量和参数的线性组合的函数可以取各种形式,如广泛使用的logit函数。

这两个传统彼此紧密关联。例如,在所有的变量——因变量和解释变量都是分类的时候,$F_i = n_i\mu_i$,式中的$n_i$是在格子$i$中的观察数。上面介绍的检验统计量也可用于拟合广义线性模型的模型评估。

## 模型的比较和选择

这个问题在分类数据分析中十分突出。研究者常常会感到有对统计模型进行比较,以及从一组模型中选出拟合较好的模型的需要。后一种情形指的是,在几个对数模型被认为是彼此嵌套时的层级对数线性拟合。这些模型有着不同的自由度,因为它们包含的参数数目不同。$L^2$中的差异给出了下结论用的比较似然比统计量。

有时研究者需要比较非等级模型。这可能是两个(或更多个)观察子群体的,或者两(或更多)个包括不同变量组的模型的模型拟合的一个比较。虽然有的时候,将这些子群体或子模型作为一个合并在一起的单独的超级模型是一种有用的处理方法,但这种路数并非总是可靠的,因为比较的那些子模型可能正好是互不相容的。因此,我们需要一种能应用于跨不同自由度的不同模型的模型拟合优度的统计标准。这一类标准中,两种使用广泛的准则是赤池信息量准则(akaike information criterion, AIC)和贝叶斯信息准则(bayesian information criterion, BIC)。

根据信息理论,AIC用一给定模型的自由度数的两倍对似然比统计量作了调整:

$$\text{AIC} = L^2 - 2(df)$$

式中的$df$表示模型的自由度数。两个模型之中,无论层级或非层级的,那个AIC值较小的模型是比较好的模型。

根据贝叶斯决策理论,BIC用总样本量($N$)对似然比统计量作进一步的调整。

$$\text{BIC} = L^2 - \log(N)(df)$$

BIC给出了用于模型比较的贝叶斯因子的对数近似值。已经证明BIC较AIC更不宜用于有较大自由度的模型,如果样本不是非常小的话。在等级模型比较也可使用$L^2$的差的时候,较之其他的拟合优度统计量,BIC通常更宜于用于简化模型。

## 分类数据分析中的缺失数据问题

任何多变量分析都非常关心缺失数据【Missing Data】问题,而对那些可能有零格的频数表模型则应予以特别的关注。尤其令人感兴趣的是所谓偏观察数据(partially observed data)。如果一个观察在所有的变量上都丢失了信息,那么这个观察根本就不能使用。但是如果一个观察只是在某些变量上丢失了信息,那么常规的列表式删除便会浪费许多有效的信息。而与这一问题随之而来的是丢失信息发生的机制。丢失数据有三种机制:完全随机地丢失、随机丢失和非随机丢失。——这最后一种在推论模型变量之间的关系时,可能会产生偏倚。近年来开发了一些用于各种丢失数据机制的分类数据统计模型的估计方法。由弗蒙特开发的LEM软件则是其中的佼佼者。

——Tim Futing Liao

(沈崇麟译校)

## 参考文献

Agresti, A. (1990). *Categorical data analysis.* New York: John Wiley.

Clogg, C. C., & Shihadeh, E. S. (1994). *Statistical*

*models for ordinal variables.* Thousand Oaks, CA: Sage.

Everitt, B. S. (1992). *The analysis of contingency tables.* London: Chapman & Hall.

Fienberg, S. E. (1980). *The analysis of cross-classified categorical data* (2nd ed.). Cambridge: MIT Press.

Liao, T. F. (1994). *Interpreting probability models: Logit, probit, and other generalized linear models.* Thousand Oaks, CA: Sage.

Long, S. (1997). *Regression models for categorical and limited dependent variables.* Thousand Oaks, CA: Sage.

Powers, D. A., & Xie, Y. (2000). *Statistical models for categorical data analysis.* New York: Academic Press.

## CATI

参见计算机辅助数据收集法【Computer-Assisted Data Collection】。

## 因果机制(Causal Mechanisms)

因果机制指某些形成结果的过程或者路径。我们通过提出一个有关一般将会出现结果的原因的假设来解释结果。正因为如此,实际上所有的社会研究的主要目的都在于发现原因。我们来考虑一个价格上涨导致消费减少的例子。联系原因和后果的因果机制是理性消费者的选择。他们观察到了价格的上涨,调整自己的消费以尽可能地扩大总效用,并降低自己对价格上涨商品的消费。总体来讲,正是这种个人层次上的理性行为造成了总消费的减少。

因果理论有两个显类:休谟理论(Humean theory),即因果如同规律(causation as regularities);因果实在论(causal realist theory),即因果如同因果机制(causation as causal mechanisms)。休谟理论认为,因果完全由可观察变量中的有关经验规律的事实构成,不存在什么潜在的因果性、因果效力或因果必然性。因果实在论则采用因果机制的观念,将因果效力视为基本的效力,并认为科学研究的任务是达到在经验上证明有关这些因果机制的理论和假设。我们来看一下各种有关"$X$ 引起 $Y$"的陈述的断言:

- $X$ 是 $Y$ 的必要和/或充分条件。
- 如果 $X$ 未曾发生,那么 $Y$ 也未曾发生。
- $Y$ 的给定 $X$ 的条件概率不等于 $Y$ 的绝对概率 $Y[P(Y|X) \neq P(Y)]$。
- 在预测【Predicting】$Y$ 值的回归【Regression】方程中 $X$ 伴有一个非零系数。
- 存在一种因 $X$ 发生而导致 $Y$ 发生的因果机制。

因果实在论的中心洞见是,最终的准则实际上是最根本的。在因果实在论看来,连接 $X$ 和 $Y$ 的潜在因果机制存在的事实解释了每一个其他的准则,其他准则都是连接 $X$ 和 $Y$ 的因果路径存在这一事实的表征。

因此,因果推论以因果机制的存在为先决条件,研究者应该努力确定连接感兴趣的变量的看不见的因果机制。下面的因果准则清单给我们提供了各种用可得到的证据检验或证实一个因果假设的方法:用相似性和差异性的密尔方法(Mill methods)作为检验必要条件和充分条件的方法;考查条件概率;考查感兴趣变量之间的相关

【Correlations】和回归;用社会理论的某些适当的部分建立有关潜在因果机制的假设。因果实在论者坚持认为,最终,经验【Empirical】证据必须要被发展到能对预设的原因和效应之间的因果机制的可信性进行评估的程度。

什么是因果机制? 因果机制是一系列由类似法则的规律支配的,从解释要素到被解释要素的一系列事件或条件。韦斯利·萨蒙(Salmon,1984)这样来阐释因果机制:"因果过程、因果交互和因果法则提供了这些机制,世界因之而运作。为了理解为什么某些事情会发生、我们需要了解它们是怎样由这些机制产生的。"(p. 132)南希·卡特赖特同样也把实在的因果机制置于她解释科学知识的核心。正如她和约翰·杜普雷所言:"事物和事件之所以有因果能力,是由于它们拥有的属性,它们有导致其他事件或状态的力量。"(Dupré & Cartwright,1988)南希·卡特赖特还从本质上对因果机制作了论证,她认为确认因果关系,需要支配正在研究的实体的因果力量或能力的实体论。因果关系不能从有关变量之间关联的事实直接推论。

构成社会因果机制基础的一般性质一直是学界争论的主题。其中可以确定的几个主要的流派是:行动者模型(agent-based models)、结构模型(structural models)和社会影响力模型(social influence models)。行动者模型采用一种将个体层次的选择结果聚合成宏观层面的成果的策略;结构模型则试图论证给定的社会结构或制度的因果效应(如征税系统)的社会成果(遵守的程度);而社会影响力模型则试图确认在主体背后影响他们的选择的因素。托马斯·谢林(Schelling,1978)给出的恰如其分的标题"微观的动机,宏观的行为",抓住了前一路数的逻辑。他的著作深刻地阐述了局部理性刻意而为的行为互动造成的,有时十分难以预测的结果。埃尔斯特对此也作了比较深刻的阐述,他认为,理性选择理论的工具支持大型社会学解释的构建(Elster, 1989)。埃米贝伊和米舍(Emirbayer & Mische, 1998)给出了当前有关代理概念争论情况的全面综述。结构主义者和社会影响路数试图确定那些社会上显著的影响(因素),如制度、国家、种族、性别和教育地位,并详细阐述了这些因素是如何影响或限制个人的人生轨迹,进而影响社会成果的。

——Daniel Little

(沈崇麟译校)

## 参考文献

Cartwright, N. (1989). *Nature's capacities and their measurement*. Oxford, UK: Oxford University Press.

Dupré, J., & Cartwright, N. (1988). Probability and causality: Why Hume and indeterminism don't mix. *Nous*, *22*, 521-536.

Elster, J. (1989). *The cement of society: A study of social order*. Cambridge, UK: Cambridge University Press.

Emirbayer, M., & Mische, A. (1998). What is agency? *American Journal of Sociology*, *103*(4), 962-1023.

Little, D. (1998). *Microfoundations, method and causation: On the philosophy of the social sciences*. New Brunswick, NJ: Transaction Publishers.

Mackie, J. L. (1974). *The cement of the universe: A study of causation*. Oxford, UK: Clarendon.

Miller, R. W. (1987). *Fact and method: Explanation, confirmation and reality in the natural and the social sciences*. Princeton, NJ: Princeton University Press.

Salmon, W. C. (1984). *Scientific explanation and the causal structure of the world*. Princeton, NJ: Princeton University Press.

Schelling, T. C. (1978). *Micromotives and macrobehavior*. New York: Norton.

## 因果建模(Causal Modeling)

在社会科学数据分析中,研究者很大程度上都依靠推断那些由测量的变量代表的概念【Concept】之间的因果关系——尤其是因变量【Dependent Variable】和自变量【Dependent Variable】之间关系的方法。尽管因果关系【Causality】问题可能是相当哲学的,但是仍然存在一些可帮助研究者理解和估计因果关系的分析路数。因果思维构成了一系列各种各样实验【Experiment】方法和基于马尔科夫链【Markov Chain】模型的基本设计基础。实验设计区别于非实验设计的重要一点是,自变量实际上是被设置为一些固定的值或被操控的,而非只是被观察。因此,用非实验的或观察的设计,因果推论要(比实验设计)弱得多。给定一个观察设计,某些条件可以加强某些事物之间,譬如说,$X$ 和 $Y$ 之间的因果,而虚假【Spurious】关系发现的论证。具有特别重要性的是时序性(即证明 $X$ 的发生先于 $Y$)。其次,也是很重要的一点是,满足因果推论假定【Assumption】的必要条件。在经典的回归【Regression】分析中,这些条件有时被叫作高斯-马尔科夫假定。

分析观察数据的方法也包括所有归入一般线性模型【General Linear Models】,或广义线性模型【Generalized Linear Models】的回归类模型,尤其是图形建模【Graphic Modeling】、路径分析【Path Analysis】或结构方程建模【Structrual Equation Modeling, SEM】,它允许多方程或联立方程系统。可以包括潜变量【Latent Variables】,以更好地表示概念【Concepts】的 SEM,可以用软件来估计。这些软件有 AMOS,EQS,LISREL 和 M-PLUS 等。

一个一般的联立方程系统可以含有双向关系【Reciprocal Relationship】(如 $X$ 引起 $Y$ 和 $Y$ 引起 $X$)。它使它们呈非递归的关系。下面是诸如这样的系统的一个例子,我们也许可以把它叫作因果模型。

$$Y_1 = b_{10} + b_{11}X_1 + b_{12}Y_2 + e_1$$
$$Y_2 = b_{20} + b_{21}Y_1 + b_{22}X_2 + e_2$$

因果模型(causal model)一词是作为联立或结构方程的同义词使用的。有时在陈述方程系统的同时,常辅以一张素描的箭头图(即一般叫作路径图【path diagram】的图表)。在估计进行之前,必须先解决识别问题【Identification Problem】。方法是通过考查外生变量【Exogenous Variable】和内生变量【Endogenous Variable】的相对数目,看一看秩和阶的条件是否得到满足。在这一因果模型中,$X$ 变量是外生的,$Y$ 变量是内生的。事实证明,在这一系统中的每一个方程都已被精确地识别时,整个系统也就被识别了,因而它的参数就可以被估计了。按照惯例,外生变量被标为 $X$,而内生变量则被标为 $Y$,但必须强调的是,如果只是简单地把它们作如是的标示,并不能使它们自动地成为外生或内生变量。相反,外生性和内生性的假定是必须被满足的。

在模型识别之后,通常都是用某些工具变量【Instrumental Variable】法作进一步的估计,如二阶最小二乘方法。此外,还有三阶最小二乘方法,但学界普遍认为它对二阶最小二乘并没有什么实质性的改进。如果系统中的每个方程都已被精确地识别,那么我们便可应用一种叫作非直接的最小二乘。在这里,我们不能使用普通最小二乘【Ordinary Least Squares】,因为它将普遍引起偏倚和前后矛盾的估计,除非系统是递归的【Recursive】,即因果性是单向的,且误差项是跨方程不相关的。有与二段和三段最

小二乘类似的最大似然【Maximum Likelihood】估计法：分别为有限信息最大似然和全信息最大似然估计法。

迄今为止已经讨论过的内容告诉我们，因果建模法必须依靠概率论作为基础的统计分析法。因果性描述一个充满不确定性的过程，而分析者的任务则是在捕捉因果模式的同时，对这种不确定性作出解释。

——Michael S. Lewis-Beck
Tim Futing Liao
（沈崇麟译校）

## 参考文献


Berry, W.(1984). *Nonrecursive causal models*. Beverly Hills, CA: Sage.

Bollen, K. A.(1989). *Structural equations with latent variables*. New York: John Wiley.

Davis, J.(1985). *The logic of causal order*. Beverly Hills, CA: Sage.

Kmenta, J.(1997). *Elements of econometrics* (2nd ed.). Ann Arbor: University of Michigan Press.

Pearl, J.(2000). *Causality: Models, reasoning, and inference*. Cambridge, UK: Cambridge University Press.


# 因果关系（Causality）

在社会科学的讨论中，当因果关系成为问题的核心时人们常常都会流露出一种犹豫之情，仿佛在发出“我们害怕因果关系”的感叹。但所有的讨论者都承认他们不能没有因果关系：“最后我们还是需要因果解释。”哲学史中因果关系的概念也遭遇了这种犹豫。因果关系的历史列车好像是一辆慢车；它就像埃希特纳赫的游行队伍，前进两步又后退一步。亚里士多德、经院学者、莱布尼兹、康德、叔本华都是因果论者。休谟、马赫、皮尔逊、赖兴巴赫、哥本哈根学派的量子物理学家们，以及罗素都是反因果论者。罗素（Russell, 1912）曾经留给我们这个大胆的说法：“我相信，因果律是逝去的时代的遗迹，就像君主制一样，只是因为被错误地当作无害的，才得以生存了下来。”（P.171）

## 亚里士多德：奠基人

我们可以把亚里士多德当作因果思想的奠基人。作为哲学家，他对各种原因做了区分，例如，质料因【causa materialis】、形式因【causa formalis】、动力因【causa efficiens】和目的因【causa finalis】。在17世纪的近代科学里，亚里士多德的动力因或劳力因（labor cause）成了最主要甚至是唯一的一个原理。于是原因就和“成果”（production）联系了起来。劳力因是主动的实施者（active agent），像一种用锤子敲打那样的外力，而效果（effect）是被动地承受，就像裂成碎片的栗子一样。这种效果的概念也体现在牛顿的外力（external power）中，它能让静止的物体运动。它还体现在社会政策制订之中。例如，19世纪法国的政策制订者启动了一个提高生育率、鼓励移民的政策，以使生育超过死亡，这样老龄化的人口结构就可以得以翻转。它还体现在行为主义【Behaviorism】学派的刺激—反应框架中。甚至在我们的语言中它也有所体现，因为下毒（poison）、治疗（cure）、镇静（calm）、加湿（humidify）和照明（illuminate）等许多动词都包含着成果的理念。

## 伽利略:必要和充分条件

也是在近代科学中,通过伽利略的努力,原因才被区分成了必要和充分的条件。按伽利略的说法,必要(necessary)意味着“必不可少的条件”(conditio sine qua non)或者是“如果没有 *X*,就没有 *Y*”,而充分(sufficient)的意思是“如果有 *X*,总是有 *Y*”。(我们应当注意,“*X* 使 *Y* 成为必需”的意思是“*X* 是 *Y* 的充分条件”。这里可能会产生混淆!)

## 大卫·休谟:历史转折点

18 世纪的大卫·休谟确实是因果关系历史的转折点。实际上,任何人今天要写这个主题的话不可能不讨论休谟。卡尔·波普尔(Popper,1972)是个很好的例子,在他的《客观的知识:一个进化论的研究》(*Objective knowledge: An evolutionary approach*)一书中,开篇第一章就是“休谟问题”。休谟不反对因果关系论(Hume, 1786)。他相信世界充满了因果关系。但他怀疑科学是否能够洞悉为何某个原因导致某个结果。休谟的出发点完全是经验主义的(empirical)。他提出,有关因果关系的知识从来不是用纯粹演绎的方法和先验的方式提出来的,而是完全基于经验。亚当不可能看到水就推论出水会淹死他。休谟还相信原因是外在于其结果的。如果一个台球和另一个球相撞,后者开始滚动,那么在第一个球上不存在任何能给我们带来有关第二个球会发生什么的提示。至于在原因和结果之间的那个中间术语(即因果之箭[the causal arrow]),休谟认为诸如成果(production)、能量、力量等都属于一种模糊的哲学,无非是迷信的庇护所和遮掩愚蠢与错误的斗篷。我们看到了火,感到了热,但我们甚至猜不出或想象不出这两者之间的联系。我们甚至不能直接察觉到意识的能量会影响我们身体的器官,致使意识的影响会经常躲过我们辛勤的考察。为什么人的意志能影响舌头和手指,而不能影响心脏和肝脏,这个问题总是让我们为难。在这个问题上,关于上帝之意志的想法让我们超出了能力的极限。我们的垂线太短了,无法探测这些巨大的峡谷。休谟的结论是,当我们同时看到原因和结果时,那么这儿就有一个“恒常联结”(constant conjunction)。过了一段时间适应了这个之后,习俗(custom)就出现了。接下来,通过某种心理联想(psychological association)机制,我们逐渐开始发展出一种信仰(belief)。而在这个信仰的基础上,我们采用因果关系的术语来进行预测(prediction)。简单说来,唯一真正存在的就是恒常联结(规律论[regularity theory]),其他的都是心理学。一言以蔽之,相关才是真问题,其余的东西都是伪问题。

## 伊曼努尔·康德:因果关系是一个先验的范畴

各路人物对大卫·休谟的回应足以装满一个图书馆。其中康德第一个承认,是休谟将他从教条主义的懵懂中唤醒,但他拒绝接受只有经验才是因果关系的基础这一观点(Kant, 1781)。他相信,先验范畴中的理解和经验一起给我们带来了那个叫作客观知识的综合体。在他看来,因果关系就是关于理解的范畴之一。这种把因果关系当作先验范畴的观点并非没有争议。约翰·麦凯(Mackie,1974)举了一个把木头劈成两半的例子,他提出,从足够远的距离来看,我们首先是看到木头的倒下,接着是听到斧子的声音,因为声和光的速度有差别,但如果我们站在跟前的话,那么就会首先听到声音,接着才看到木头分开倒下,因为物体有惯性。如果桌子上的小火车撞上了另一辆火车并使得后者开始运动起来,那么这跟我们

最初的意识机能并没有什么关系，因为桌子底下可能藏着磁铁，引起了第二辆火车的运动。因此，与其说这里并没有太多的先验的因果知识，毋宁说是在尝试着对知识进行假定，为了得出一个结论我们需要经验和观察。

## 现代科学：假设检验

我们今天的科学研究中经常用到的假设检验的传统是由奥古斯特·孔德、约翰·斯图加特·密尔等19世纪实证主义哲学家提出的，进而在20世纪的实证主义研究中成了标准的程序。卡尔·波普尔（Popper，1959）和卡尔·亨普尔（Hempel，1965）提出的演绎律则法（deductive-nomological approach）也对这种基本态度有所贡献，演绎律则法通常被称为覆盖律理论（the covering law theory）（即每个具体的因果陈述都为普通法则所覆盖，普通法则作为一般性的理论和假设在背景中起作用）。

## 因果关系的概率论

与此同时，对于因果关系的争论还在继续。大多数提法都有一种包围策略，因为因果成果并不是以直接的方式达到的，而是通过一些其他的标准间接达到的。这些标准之一就是"概率（probability）"，有一派也许可以叫作因果关系的概率论（a probabilistic theory of causality）的信徒的科学家就使用这种概率。"因果关系的概率论"也是一本书的名字，其作者帕特里克·祖佩在1970年代开创了这一派的思想。他的观点符合世界上大多数科学研究者的观点——他的观点是，当且仅当 $X$ 存在（$X$ 的概率大于0），$X$ 在时间上先于 $Y$（$t_X$ 这一刻先于 $t_Y$ 到来），在 $X$ 和 $Y$ 之间存在统计关系（在给定 $X$ 条件下，$Y$ 出现的概率大于 $Y$ 独自出现的概率），且不存在伪关系（当控制其他潜在的混杂【Confounding】因子时，$X$ 和 $Y$ 的统计关系不会消失），$X$ 才是 $Y$ 的一个原因。这些因果关系的标准被用于社会科学，特别是统计分析之中。休厄乐·赖特1920年在遗传学中所使用的路径系数法、格尔曼·沃尔德1950年代在计量经济学中使用的联立方程模型、西蒙和布莱洛克1960和1970年代在社会学和其他社会科学中所使用的因果模型，以及卡尔·约雷斯括格1970年代及以后的线性结构关系系统（LISREL），只不过是许多例子中的一小部分而已。这一派思想始终要面对三个问题：第一个是概率的概念，它解释的是统计关系（也即相关），而非因果关系；第二个是对理论的依赖，这体现在因果模型中的变量关系上，尤其是控制虚假关系的那些因子上；第三个问题使用的是时间优先，而非原因优先，也就意味着研究实践建立在"在它之后，也正是由于它"（post hoc，ergo propter hoc，即，此后，所以因此）这个诡辩之上。

## 马里奥·邦格：简单性的天堂

另一种包围策略是采用伽利略的"必要"和"充分"条件准则，许多当代作者都这么做，但用得最聪明的是约翰·麦凯（Mackie，1974）。他反对马里奥·邦格（Bunge，1959）的极端立场。马里奥·邦格告诉我们，如果我们把因果关系的概念搞得太宽泛，我们就是犯了意义通胀（meaning inflation）罪。在邦格看来，事实上有许多形式的确定（determination）——概率的、辩证的、函数的、结构的、机械的和其他形式的确定，而因果关系形式的确定只是其中一种而已。这种因果关系的确定的核心是亚里士多德所给出的因效（causa efficiens）（即必要的成果）准则。除了成效（productive）之外，因果关系还是条件的（如果 $X$，那么 $Y$），唯一的（一个原因，一个结果），不对称的（当 $X$

引起 $Y$ 时，而 $Y$ 却不会引起 $X$），恒久的（没有例外、非概率的因果关系）。邦格还给出了其他的严格准则，例如线性、可加性、连续性，等等。他的因果关系的概念非常严格。他反对功能论者、互动论者、辩证论者，因为辩证论者用相互依赖来看待一切事物，因此他们也是浪漫主义者。在他们看来，因果关系是打开每一扇门的钥匙，是一帖灵丹妙药。邦格认为，如果严格定义的话，因果关系只是现实的一小部分，但却是确实存在的。因果关系既非神话，也非灵丹妙药。

## 约翰·麦凯：因果场中的 INUS

不言而喻，邦格（Bunge，1959）对因果关系的严格定义是一个简单的天堂。麦凯（Mackie，1974）反对这种严格，发展出了一个非常宽泛的、接近日常语言的因果关系概念，它兼容概率论、多因论、目的论、功能论以及其他的确定形式。他的方法的依据是充分条件和必要条件。在他看来，一个起因就是一个“在这种情况下”的必要条件，也就是说，那些沉默的前提和细节应该都尽可能说清楚。例如，在一幢房子里发生了火灾。专家称这是短路引起的。他们没有给出一个必要条件，因为除了短路之外的其他事情，例如电热油汀翻倒等情况也可能会引发火灾。他们没有给出一个充分的条件，因为即使出现短路，其他的条件，例如可燃材料或者缺乏有效的灭火器等，也是火灾发生所必要的。因此，它是一个复杂的条件集合，有积极的条件也有消极的条件，与短路一起为火灾构成了充分但并非必要的条件，因为其他的因素也可能会引发火灾。短路只是这个条件集合中的一个必要的部分，因为如果没有短路，可燃材料或者缺乏有效的灭火器无法引起火灾。这样，短路就是复杂的条件集合中一个必要但不充分的部分，而这些复杂条件集合，对于结果而言，则不是必要的，但却是充分的。简言之，短路对于起火来说是一个 INUS 条件——一个集合中不充分但是必要的部分，这个集合对于结果来说是非必要的，但却是充分的。麦凯在这一点之后还添加了一点：因果场中总是有一些因素是不变的、固定的。例如，从严格的逻辑意义来说，出生就是死亡的一个 INUS 条件，但它却不能算作死亡的一个候选的原因，因为对死因的声明会指涉那些曾经活过的人们。另一个例子没有这么明白，这是斯里文（Scriven，1959）有关梅毒的例子。梅毒螺旋体（treponema pallidum）是一种病毒，它是梅毒唯一的病因。但是，感染梅毒螺旋体病毒的人中只有很少比例的人会发展到最后的第三期，全身麻痹（paralysis generalis），这时候会因大脑瘫痪、肌肉运动失调和精神失常而死亡。现在，有关唯一病因的陈述指的是一个不同的因果场，而非第二个关于全身麻痹的陈述。第一个是所有易感染这种病毒的人（比方说所有曾经有过性交的人）的因果场。第二个是已经感染了这种病毒的所有人的因果场。

在研究实践中，这种因果场的概念特别重要，因为它还包含那些差不多可以算是不言自明的与研究问题相关的确定因素，这些因素之所以被确定，是出于时间和空间的（生活在一个国家，在此国家作研究）实际考虑。这些因素在 INUS 条件之外，但因存在异质性的危险而被确定（在失业的因果场作研究，但却遗漏了青年工人，因为青年失业是一个完全不同的问题）。在社会科学中，最大的过错是事实上忽略了因果场。

——Jacques Tacq

（杨可译　沈崇麟校）

## 参考文献

Aristotle. (1977). *Metaphysica*. Baarn: Het Wereldvenster.

Bunge, M. (1959). *Causality: The place of the causal*

*principle in modern science*. Cambridge, MA: Harvard University Press.

Hempel, C. G. (1965). *Aspects of scientific explanation and other essays in the philosophy of science*. London: Collier-Macmillan.

Hume, D. (1786). *Treatise on human nature*. Oxford, UK: Claredon.

Kant, I. (1781). *Kritik der reinen Vernunft* [Critique of pure reason]. Hamburg: Felix Meiner.

Mackie, J. (1974). *The cement of the universe*. Oxford, UK: Oxford University Press.

Popper, K. (1959). *The logic of scientific discovery*. London: Hutchinson.

Popper, K. (1972). *Objective knowledge: An evolutionary approach*. Oxford, UK: Clarendon.

Russell, B. (1912). *Mysticism and logic*. London: Allen & Unwin.

Scriven, M. (1959). Explanation and prediction in evolutionary theory. *Science*, *130*, 477-482.

Suppes, P. (1970). *A probabilistic theory of causality*. Amsterdam: North Holland.

Tacq, J. J. A. (1984). *Causaliteit in Sociologisch Onderzoek. Een Beoordeling van Causale Analysetechnieken in het Licht van Wijsgerige Opvattingen over Causaliteit* [Causality in sociological research: An evaluation of causal techniques of analysis in the light of philosophical theories of causality]. Deventer, The Netherlands: Van Loghum Slaterus.

Wright, S. (1934). The method of path coefficients. *Annals of Mathematical Statistics*, *5*, 161-215.

## CCA

参见典型相关分析【Canonical Correlation Analysis】。

## 天花板效应(Ceiling Effect)

在一个量度拥有一个潜在的应答(responses)的明显上限,并且参与者的得分大量集中在这一上限或这一上限的附近时,天花板效应就发生了(它的反面是地板效应【Floor Effect】)。尺度的弱化(scale attenuation)是一个方法学上的问题。每当变异以这种方式受到限制时,尺度弱化问题就会发生。虽然通常这个问题都在实验研究的语境中进行讨论,但是当一个结果的量度被证明在它的潜在值域的上限几乎没有变化时,那么,这个问题就会对任何类型研究的信度都有威胁。

天花板效应可能由各种因素引起。如果结果量度是某种有上限的表现性评价任务(performance task)(如若干正确的答案),参与者可能发现这个任务太容易了,因而在这个量度上的得分几乎都是完美的。等级尺度也有上限,而参与者的回答可能落在应答尺度的上域之外。

由于缺乏变异,天花板效应就会对实验和非实验的研究效度产生严重威胁,因而在对任何有这种效应的研究结果进行解释时,都必须加倍小心。如果一个因变量受到了天花板效应的影响,那么实验和准实验的结果就会显示无效应。例如,在一个高分意味着赞许,但最高的答案却未能捕捉到最为赞许的评价时,那么在这样的态度量表中就有可能发生天花板效应。因为这个量度未能评定更高的赞许态度,如果参与者先前的态度已经在态度分布的高端,那么用这一量度

进行的说服性研究(persuasion study)就可能产生零发现。

在解释交互效应【Interaction Effect】的时候,天花板效应尤其麻烦。许多交互效应都会遇到在一个自变量水平没有发现显著的差别,而在另一个水平上却发现了显著的差别。没有发现显著差别的原因可能是天花板效应,因而对诸如这样的交互效应的解释应该加倍小心。

天花板效应也威胁到非实验或相关研究。大多数推论统计都基于所涉及变量的正态分布的假定。天花板效应可能会导致无法进行显著性检验,或因为方差的值域受到限制而加大在虚无假设实际为真的时候被拒绝的可能性(第二类误差【Type Ⅱ Error】)。

天花板效应问题的最佳解决办法是进行前测(pilot testing),它可以使我们尽早发现问题。如果发现了天花板效应,我们可以有几种方法来处理。如果结果量度是任务绩效,那么我们就可以使任务变得更为困难,加大潜在答案的值域。使用等级尺度就可扩大潜在的应答数。在等级尺度中,还有另一种解决方法,这种方法涉及等级尺度中极端锚刺激(extreme anchor stimuli)的使用。在对感兴趣的变量进行评定之前,请参与者对一个刺激评级,要求他们使用分布的极端。这种做法将鼓励参与者使用等级量表的中间值域,从而减少天花板效应发生的可能性。

——Robert M. Hessling
Nicole M. Traxel
Tara J. Schmidt
(沈崇麟译校)

## 格(Cell)

格这一词是指任何一张列联表【Contingency Table】中的行与列的交叉点。因此,一个有两个二分变量的列联表将会有四个格。

——Alan Bryman
(沈崇麟译校)

## 截删数据(Censored Data)

在以下四个条件都得到满足的时候,统计数据文件就被认为是截删的:(a)得到的数据是来自一个总体【Population】的概率样本【Probability Sample】;(b)因变量有一或多个极限值;(c)对于所有的个案,不论因变量【Dependent Variable】的值是什么,都有可用自变量【Independent Variable】的值;(d)极限值并非因为个案选择了这个值。所谓极限值是这样一个因变量的值,即无论是低于还是高于这一值都没有其他被量度或观察的值。举一个下限的例子,如街区的犯罪率是不能小于零的。在有零犯罪率的街区,可能有两种在“犯罪倾向”上完全不同的街区类型,第一种类型是非常安全的社区,在那里没有发生过犯罪。第二类街区则有若干“易于导致犯罪”的特性,只不过碰巧在一个特定的年份没有发生犯罪。尽管这些街区之间存在着差异,但这两个街区都有零犯罪率作为它们的下限,于是就不存在任何的因变量值可把这些街区彼此区分开

来。各种类型的售罄事件都是作为总需求量度的销售上限的例子。有些（售罄）事件较之其他事件可能有更高的未满足的需求，但是我们却无从对此作出判断。

布林（Breen，1996）把有极限值的因变量叫作删节（truncated）。如果个案没有因为它们在因变量或自变量的值而被排除在研究样本之外，那么总的来讲，数据就满足了被叫作截删数据的第三个条件。在这个意义上讲，因变量的量度无论低于下限，或高于上限都是不可能的，因此是被截删的。就街区犯罪问题而言，如果城市中所有的街区，不论它们的犯罪率高低，也不论它们的自变量的值是什么，都包含在了样本中，那么这个数据就是截删的。如果研究根据某一变量的值排除了一些街区，那么数据便是删节（trunction）的，而不是截删的。数据是适用于高犯罪率或低犯罪率街区，或者只用于少数民族或非少数民族的街区，则被称作删节的，因为数据收集止于变量的某个值，或街区的某个特性。

为了满足截删的第四个条件，有极限值的个案一定不能是个案自身自我选择的结果，就像在个体数据上可能发生的那样。正如布林（Breen，1996）所言，那些没有购买奢侈品的家庭选择了或选定了自己不做奢侈品的购买者，但不论街区的居民可能采取了多少安全防范措施，街区本身都决定不了街区内是否将会发生犯罪。就购买这一例子而言，购买数据分析可能需要的是样本的选择，而非截删数据的方法。

——Dannis W. Roncek

（沈崇麟译校）

＊也可参见截删和删节【Censoring and Truncation】。

## 参考文献

Breen, R. (1996). *Regression models: Censored, sample selected, or truncated data.* Thousand Oaks, CA: Sage.

# 截删和删节（Censoring and Truncation）

截删和删节二者都涉及随机变量的经验数据是不完整或部分的这样一种情况。纵观整个现存的文献，这两个词的使用其含义并不那么一致，有些作者称之为截删，另一些作者则称之为删节。有鉴于此，一个作者在给这两个词的其中一个定义时需要慎之又慎。

一般来讲，截删涉及这样一种情况，随机变量的某些值因为位于某个值域内而被记录，所以量度是不够精确的，至少对样本的某些成员而言，量度是不精确的。与之不同，删节在习惯上是指在随机变量的值在某一值域内时，某些总体成员的信息未被记录下来这样一种情况。由此可见，截删与不完全或不精确的量度有关，而删节则与不完整的样本相关，即样本的选取是以随机变量的值为条件的。哈尔德（Hald，1952）是最初区别截删和删节的作者之一。自那时起至今，分析截删变量和删节样本的方法有了很大的发展（Schneider，1986）。

截删变量是一种因为变量的某些值位于某个或某些值域内而使样本中的某些观察未被精确量度的变量。截删观察或个案则指特定变量被截删了的样本成员（即只知道位于某一值域内）。

例如，在1990年美国人口普查中，每个报告自己实际年龄在90岁或90岁以上的人，其年龄在公开发表时都在90岁作截删

（都按 90 岁处理）。美国普查局之所以在 90 岁作截删，一是为了保密，二是为了避免不精确的信息（因为老人有时会夸大自己的年龄）。凡是在家庭表中报告其年龄在 90~120 岁的那些人，出现在宏观数据中，一律作为年龄在 90 岁截删的个案。

删节变量是一种，如果变量的值位于某一或某些值域内，观察才被量度而成为样本成员的变量。与之等价的是，如果变量的值不在选定的值域内，那么这一变量就根本不会被量度。从概念上讲，删节样本类似于一种将疑似异常值的极端值去除后的修整过的样本。删节和修整之间的一个差别是，样本的删节习惯上以变量的数值为依据，而修整则以变量的有序统计值为依据（例如，将上端和底端的 5%排除在外，尽管它们在数据中是存在的）。例如，个子很矮的人都被排除在兵役之外。正因为如此，军事人员的样本身高是一个删节变量，因为如果人们的身高低于某一水平，那么总人口中的这些人便将系统地被排除在军人之外。

## 截删的类型

有几种不同的方法对截删和被截删变量进行分类。一种基本的区别是随机截删和基于某些预先确定的准则的截删。两种比较普遍的预先确定用于截删变量的准则叫第一类和第二类截删（有为数不多的作者也使用第三和第四类截删，但其他类型的定义，至今仍仁者见仁，智者见智，颇多歧义）。

由因变量的值落在了某个外部决定的值之上或之下而引起的截删，叫作基于长度的或第一类截删（length-based censoring），当美国普查局，像上面介绍的那样以 90 岁为准，对人们的年龄进行截删时，这样的截删就是基于长度的截删。

当变量的样本值是有序的（通常是从最低值到最高值），且在事先确定的那部分有序样本值记录完毕之后，便停止进行精确的量度时，便会产生第二类截删。第二类截删主要发生在变量涉及事件的时间（如一个事件的时间，例如死亡），在预先选定那部分样本的事件时间已经记录完毕之后。其余的样本成员只知道有超过量度终止的那个时间（即大于最后的观察时间）。例如，社会科学家可能记录 1 000 个因工厂倒闭而刚刚失业的失业时间。如果研究者事先决定样本成员的 60%发现新的工作之后，便不再继续记录失业延续的时间，那么他们就是根据失业延续的时间在进行第二类截删。截删个案部分的比例被事先固定在了 40%，而观察的失业延续的最长时间取决于样本 1 000 个人中的 600 个找到新工作的时间的长度。

截删也经常会是其他的随机过程的结果。为了避免有偏误的结论，感兴趣的随机变量必须在统计上独立于引起截删的随机过程。如果二者是在统计上独立的，那么，截删就可以说是非信息型的；反之，如果它们在统计上是相关的，那么截删就是信息型的。

例如，假设感兴趣的变量是妇女生育第一个孩子时的年龄。有些妇女至死没有生过孩子，那么对这些妇女而言，她们的第一个孩子出生年龄的截删是在她们去世时候的年龄。非常清楚，死亡是一个随机过程，而问题的关键在于妇女死亡的年龄是否独立于她生育第一个孩子的年龄。我们没有统计上的根据来确定，这一过程是否会产生有兴趣的结果，以及这一随机过程是否在统计上是独立的。因此，有关它们的独立性的决定在实践中将根据实质性的论据作出。在任何给定的应用中，有些学者可能论证这两个过程在统计上是独立的，而有些学者则认为它们是不独立的。不仅如此，还有一些别的学者，认为这两个过程在某些条件下是独立的（对所有实际的目的），但在其他条

件下是不独立的。例如，控制妇女生育第一个孩子的过程和引起妇女死亡的过程，在像瑞士这样的发达国家（几乎）是独立的，而在撒哈拉沙漠以南地区的不发达国家是不独立的。

右、左、双重和多重截删也是一种区分截删模式的方法。右截删的意思是，如果变量的值超过某个水平$\tau_R$，那么这些变量则未被作精确的观察；左截删的意思是，如果变量之值小于某个水平$\tau_L$，那么这些变量便未被作精确的观察。双重截删则指左右截删的组合，它的意思是在变量之值高于某个水平$\tau_L$，低于某个水平$\tau_R$，也即在$\tau_L$和$\tau_R$之间的那些值便会得到观察。最后，多重截删则发生在有多个截删水平，且数据只记录变量值落入的区间。

例如，个人收益可能在 100 000 美元处右截删（即高于 100 000 美元，在 100 000 美元处被截删），或在 5 000 美元处左截删（即低于 5 000 美元，在 5 000 美元处截删）。此外，也可以只记录 5 000~100 000美元的值，而把两种极端值都作截删处理，于是就有了双重截删。最后，我们不难发现，有些调查要求人们报告他们的年收益，处在几个有序的值域中的那一个（如小于 5 000 美元，5 000~10 000 美元，10 000 ~ 20 000 美元，……，90 000~100 000美元，100 000 美元以上）。那时，答案就是被多重截删的，因为被记录的是一个定序变量，而非一个原始的连续变量。

## 截删结果分析

托宾（Tobin，1958）提出的消费支出截删正态回归模型，是最早将截删理念包含在经验社会科学研究中的截删结果分析方法之一。在截删正态回归模型中，真因变量$Y^*$是在某个水平$\tau_L$之上（和/或在某个其他水平$\tau_R$之下）被充分观察的。否则，就只知道因变量的值低于水平$\tau_L$（或高于$\tau_R$，或在$\tau_L$到$\tau_R$这一值域之外）。这就是说，连续变量$Y$的值可能是左截删、右截删，或左和右截删的。这三种情况可作如下正式的表达：

1.左截删：

$$\begin{aligned} Y &= Y^*,\ \text{如果}\ Y^* \geqslant \tau_L \\ &= \tau_L,\ \text{如果}\ Y^* < \tau_L \end{aligned}$$

2.右截删：

$$\begin{aligned} Y &= Y^*,\ \text{如果}\ Y^* \leqslant \tau_R \\ &= \tau_R,\ \text{如果}\ Y^* > \tau_R \end{aligned}$$

3.左和右截删：

$$\begin{aligned} Y &= Y^*,\ \text{如果}\ \tau_L \leqslant Y^* \leqslant \tau_R \\ &= \tau_L,\ \text{如果}\ Y^* < \tau_L \\ &= \tau_R,\ \text{如果}\ Y^* > \tau_R \end{aligned}$$

我们假定$Y^*$有累积概率分布函数$F(Y^*)$。如果$Y^*$有正态分布，且均值$Y^* = X'\beta$，那么这个模型便叫作 Tobit 模型。这个模型很容易用最大似然法进行估计。其他截删回归模型也可以立即用最大似然法进行估计，只要$Y^*$的累积概率分布是可以被设定的。

如果真因变量$Y^*$是在$n$个点上多重区间截删，有$\tau_1<\tau_2<\tau_3< \cdots <\tau_n$，且$F(Y^*)$有某种已知的或公设的分布，那么定序变量的回归模型也可以用最大似然进行估计。如果$F(Y^*)$有累积正态分布，那么这个模型便叫作定序 probit 模型；如果$F(Y^*)$有累积的 logistic 分布，那么这个模型便叫作定序 logit 模型（Maddala，1983）。

在托宾提出截删正态回归模型的同一年，卡普兰和迈耶（Kaplan & Meier，1958）提出了在$T$、$F(t)$的分布未曾事先设定的，一个事件的随机时间长度$T$（例如一直到死亡的时间）的右截删数据的$(1-F(t))$的极限

估计量(product-limit estimator)。他们提出的估计量,现在叫作 Kaplan-Meier 估计量,它是无偏和渐近正态的。它在生物统计学、工程和社会科学中有关事件发生时间,如死亡或失效时间的截删数据的分析中的应用呈爆发式增长。

## 删节

与截删一样,样本可能在左、右或双重删节,在总体成员因为变量之值小于$\tau_L$而被排除在样本之外时,样本就是被左删节的。在总体成员因为变量之值大于$\tau_R$而被排除在样本之外时,样本就是被右删节的。在总体成员因变量之值小于$\tau_L$,或变量之值大于$\tau_R$而被排除在样本之外时(即值不在$\tau_L$和$\tau_R$之间),样本便是被双重删节的。在某些场合,$\tau_L$和$\tau_R$是预知的,但在另一些场合,$\tau_L$和$\tau_R$是未知的,它们是与其他的随机变量的样本值一起被估计的。

分析删节数据几乎总是比分析截删数据的问题要来得多。在数据被截删时,我们知道数据哪些部分作了完整的量度,哪些部分未作完整的量度,同时也知道截删点的值。在分析删节数据时,整个总体的真的变量分布是颇费思索的。此外,由于不知道被删部分在目标变量上的取值,人们也不知道总体中被删去的是哪一部分。然而,如果变量的真概率分布是已知的,那么依据它究竟是左、右还是双重删节的,便可知它是大于$\tau_L$,小于$\tau_R$,还是在$\tau_L$和$\tau_R$之间的,我们便可估计一个结果的条件分布模型。

瓦克泰和特拉塞尔(Wachter & Trussell,1982)用18世纪英国征募海军的历史数据,讨论了英国人身高的估计问题。那时,征募的海军新兵对身高有一定的要求。这一要求意味着海军新兵是18世纪英国人的一个删节样本。在他们的应用中,他们假定身高有正态分布,因为在其他一些非删节的样本中,正态分布和人的身高数据拟合得很好。他们也处理了一些其他的异乎寻常的问题,如在使用的删节水平中的随机波动、在某些值上的堆积、凑整(rounding)等。

——Nancy Brandon Tuma

(沈崇麟译校)

## 参考文献

Hald, A. (1952) *Statistical theory with engineering applications*. New York: John Wiley.

Kaplan, E. L., & Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, *53*, 457-481.

Maddala, G. S. (1983). *Limited-dependent and qualitative variables in econometrics*. Cambridge, UK: Cambridge University Press.

Schneider, H. (1986). *Truncated and censored samples from normal populations*. New York: Marcel Dekker.

Tobin, J. (1958). Estimation of relationships for limited dependent variables. *Econometrica*, *26*, 24-36.

Wachter, K. W., & Trussell, J. (1982). Estimating historical heights. *Journal of the American Statistical Association*, *77*(378), 279-293.

# 普查(Census)

普查是一种对整个统计总体【Population】进行计数的调查。尽管这个词可用于许多不同类型总体(如交通普查,计数在一个特定的时期,通过一个特定地点的

全部车辆),但它一般,以及在本词条中则特指人口普查——一种人口学上的全体调查。它涉及计数一个特定的地理区域,通常是一个国家,在一个短时间内的全部人口。普查既是横截面【Cross-Sectional】的(它们提供单独一个时间点的结果),也是纵向【Longitudinal】的(它们提供一系列有规律的计数,使我们能在时间上追踪一些地点的特征)。

普查是世界上进行得最普遍的社会研究调查。联合国通过传播上佳的实践指南、定义和出版世界上大多数国家的人口统计资料来协调各国的普查工作。

在普查中收集的最低限度的信息涉及人口的年龄、性别和家庭的组成。而更进一步的个人特征的收集一般则包括种族、就业状况和学历等。此外,许多国家也收集住房方面的情况。还有些国家,普查还要调查宗教派别,或上班路途的详情。财务信息,特别是收入方面的调查在世界各国的普查中差异很大,因为在有些文化中,人们不愿意公开诸如这样的信息。

普查的组织工作面临巨大的挑战。它要在调查的实地组织足够多的调查人员,对他们进行充分的培训。与此同时,还必须进行充分的协调安排,以确保在普查日,或至少在当晚,调查的对象身在调查的地点。诸如这样的普查工作人员需要根据地理框架来组织,以确保每一个居住地(所谓居住地,除了家庭户、社区建筑和医院之外,还必须包括无家可归者过夜的街道、露天居住地或人们在普查之夜工作的地点)等。

为了高效地收集普查数据,我们必须有需要进行信息收集的地方(调查地)的一份完整的清单。每一个居住地,以及每一个居住地居民都必须只被统计一次,如果被统计了两次(如学生可能既在自己的家中被统计,又在所在的大学被统计),那么就必须加注,并在总人口统计中对之作适当的调整。这一目标可以通过掌握一个普查的地理框架来达到。

使用最多的框架是以地图为依据的,该地图比例尺要足够大,以能显示每一可能在使用中的楼房。在那些没有精确测绘地图的国家,可以使用高精度的卫星或空中摄影。地图要被分成几个互不重叠的区域,每一个区域由单独一名人口普查员或单独一个人口普查组负责。这些地区必须完全填满这张地图。确保每一个需要涵盖的地点都被包括在内,每一个需要访查的地方只被分配一次访查。这些——“列举”的区域,构成了访查“投入”的地理。每一个列举区域的普查结果也都会公开出版。现代普查也同时记录每个访查过的住户或建筑的精确的地图参照系(纬度和经度或坐标系)。参照系的数值既可以从地图上,也可以从全球位置测定系统(GPS)得到。

收集诸如这样的地图参照系的长处在于,它可以知道每一个普查地的地理编码,这使我们非常容易就为任何一个想要“输出结果”的地理区域计算普查数字(一般使用的都是地理信息系统)。这允许我们出版的普查数据涵盖的单位,不同于“输入”的单位。而输入单位的定义则需要考虑能使数据收集的团队进行有效管理,而非普查数据用户的需求。

基于地图的普查地理系统的备择系统是基于地址清单和与之关联的街道地图系统。这些系统需要在普查前制作一个可靠的“基本地址文档”。这个文档必须完整,并容易和一张地图关联。这种方法的一个很好的例子是美国使用的 TIGER(topologically integrated geographic encoding and referencing,地志综合地理编码和参照)系统。这一系统现在因使用基本地址文档(master address file,MAF)而得到了增强,它使我们可邮寄,而非由普查员投递普查表。用地址文档的困难之处在于,我们必须保持

这些文档的内容足够新，这样它才能给普查提供可靠的依据。列举处理使这样的文档有效而完备。

普查需要尽可能精确，因为它的结果将用于重要的法律、行政管理和研究。在美国，执行一个简单的人头统计的普查已载入宪法，因此国会议员的选区的边界可能因此而重划。这种做法保证了基于等人口地选出每一个代表的公平选举。普查的结果使中央政府的资金能根据地方的人口多少和人口的需要分配到地方。普查通过提供"分母"给大量其他社会研究调查提供数量基础。这就是一个地区的基本人口数字，这样就可以据此计算各种率。例如，一个地区有多少失业者在申请福利这样的数字意义不大，除非能估计这一统计数字占经济活动人口的多少百分比。普查提供的那些基本数字，使我们可以计算百分比失业率，并进行可靠的地区之间的比较。

虽然普查是一种得到在国际上可比较的人口估计值的方法，但是它们正在变得越来越难执行，特别是在城市地区，因为人口流动和难以统计人口的存在，情况尤其如此。普查侵犯了填表人的隐私【Privacy】，虽然大多数公民或认为这是政府的职责所在，或因为不愿填表可能受到处罚而接受这样的侵犯，但是不管怎么说，普查是越来越难执行了。普查机构对普查数据采用了最高的保密标准，以确保提供了隐私数据的人口的安全，保证他们提供的数据不被滥用。虽然如此，对人口中某一部分进行统计，还是变得越来越困难。例如，那些在法律地位上存在问题的移民就不愿意被调查。在某些国家也有类似的情况，单身的年轻人都比较难调查，因为在普查员打电话进行联系时，他们常常没有待在家里，且他们也对表格的填写缺乏兴趣。

因为进行普查的费用和困难的增加，许多国家改而进行调查，推行备择的对人口进行追踪的登记系统。诸如这样的登记，既可以是特定的人口登记，要求公民登记住址或家庭组成发生的任何变化；也可以是其他例行进行的信息收集，如健康登记、选举登记、财产税明细或驾照等。有情况表明，似乎在不久的将来，登记系统将有可能在大多数发达国家变成一种惯例。而普查或"样本普查"（更加确切地讲，它是以从总体抽取的样本为根据的"人口调查"），则被用于保证登记系统数据的精确性。这种方法随时保持对人口进行更好的追踪，有着每 10 年一次，或即使每 5 年进行的人口普查所无法比拟的长处。

普查仍然是许多国家人口学研究的"黄金标准"，并通常都是单一的最重要的社会数据的来源。然而随着知识社会的到来，信息技术几乎无处不在，来自许多源头的有关个人的信息可以可靠而合法地合并，这将使得普查渐渐式微。到那时，它将是人口科学的基础，并提供为许多其他社会研究更为倚重的基础数据。

——Robert Barr

（沈崇麟译校）

## 参考文献


Barr, R.(1996). A comparison of aspects of the US and UK censuses of population. *Transactions in GIS*, *1*, 49-60.

Dale, A., & Marsh, C.(Eds.).(1993). *The 1991 census user's guide*. London: HMSO.

National Research Council.(2001). *The 2000 census: Interim assessment*. Washington, DC: National Academy Press.

Rees, P., Martin, D., & Williamson, P.(Eds.).(2002). *The census data system*. Chichester, UK: Wiley.

United Nations Statistics Division.(2001). *Handbook on census management for population and housing censuses-Series F* (No. 83, Revision 1). New York: United Nations.

## 普查调整(Census Adjustment)

美国的普查试图每10年一个一个街区(街区是普查中最小的地理单位,街区的面积随人口密度的变化而变化,全美大约有700万个街区)地计算所有美国的居民。州和亚州(substate)的人口统计旨在为众议院的席位分配、联邦资金的分配、国会选区的重划和市政规划等提供依据。计数人口是很困难的,常会发生两种误差【Error】:总数遗漏(gross omissions, GO)和错误计数(erroneous enumerations, EE)。GO是未能计数应该计数的人所致,而EE则是因为对人作了错误的计数。在一个错误街区计数一个人,则会同时引起GO和EE。一般来说,GO的数目会稍大于EE,因而造成在人口和地理上的不均匀的少计。1980年、1990年和2000年,美国普查局试图用一种基于捕捉-再捕捉法的双系统估计(dual-systems estimation, DSE)对普查的统计数字作调整,以降低有差别的少计(differential undercount),不过尝试没有成功(有些国家则用其他的方法对普查数字进行调整)。有关这方面的讨论请参阅《美国调查方法》(*Survey Methodology*, 1992)特刊、《美国统计协会杂志》(*Journal of the American Statistical Association*, 1993)、《统计学》(*Statistical Science*, 1994)和《社会》(*Society*, 2001)。DSE涉及以下几个方面:

- 抽取一个街区的随机样本(2000年,DSE抽取了25 000个街区)。
- 在普查的次日,进行一个普查后的调查(post-enumeration survey, PES);试计算这些街区的居民。
- 试根据那些常常是不完整或不正确的数据,把PES记录与普查记录数据相配。
- 在被称作后分层的(post strata)(2000年,DSE用了448个后分层)人口组内,通过将捕捉-再捕捉估计值与普查数字比较,估计少计数。

在普查中进行的计数叫作捕捉,在PES中进行的计数则叫作再捕捉。PES试图确认EE数目,并对移动者(mover)作出解释。最终这将产生每一个后分层的调整因子,即将后分层的估计人口数除以后分层的普查的人口统计数。而美国每一个街区的普查人口数的调整是这样进行的,先一个层一个层地将街区的普查人口统计数分隔成它的组成部分,再将每一层的调整因子应用于它对应的普查人口数,然后将街区内所有层的调整人口数加总,便可得到街区的调整总人口数。这一方法的依据是这样一个综合假定(synthetic assumption):少计率在每一个后分层内都是一个常数,不论它们地理位置如何。若综合假定不能成立则称为异质性【Heterogeneity】。

还有另一种估计人口数的方法,叫作人口分析(demographic analysis, DA)。DA用行政记录及迁进和迁出的居民的估计数来估计人口数。估计使用的是如下的恒等式:

$$人口数 = 出生数 - 死亡数 + 迁进数 - 迁出数$$

1980年,美国普查局决定不再对普查结果进行调整,因为丢失的数据太多。1990年,普查局力图对普查数字进行调整,但是行政管理部门却推翻了普查局的意见,因为他们认为调整不可能提高普查数字的精确性。普查局打算调整2000年的普查数字,但最终还是决定不再作调整,因为DSE的结果与DA相左。根据2000年的普查,大约有2亿8 140万人居住在美国。经调整

之后,增加了 1.2%,但是 DA 却发现普查人口比实际多 0.7%。

美国的普查是非常精确的,建议要作出的调整是比较小的。例如,在 2000 年,德克萨斯的人口比例从 7.409 4% 调到了 7.452 4%,增加了 0.043%。而这已经是比例改变最大的州了。为了改进普查数据,对这样微小的改变调整必须极为精确。在普查中或 PES 中的回答误差将会给记录的相配带来问题,并产生 DSE 相关尺度上的较大处理误差(processing errors)。异类和相关【Correlation】偏倚也会产生较大的误差。

DSE 始于捕捉-再捕捉,但复杂的层上层,其细节对总体估计有很大的影响。例如,有的程序为了得到数据,通过代理访谈,在相邻的街区搜索丢失的记录,探测重复记录,以找出那些在普查日和 PES 之间迁移的人,并输入丢失数据。DSE 的关键在于将 PES 记录与普查记录精确地相配,独立性和综合性地假设,而不是更好地计数。

DSE 要在 2000 年的普查人口数上净增 330 万。截至 2001 年 10 月,普查局估计净处理误差在 500 万~600 万,而 DSE 的粗误差则在 1 200 万以上。相比之下,粗普查误差估计在 1 000 万左右(U. S. Census Bureau, 2001),调整的误差至少与 DSE 打算要纠正的普查误差一样大:调整美国的普查数据可能很容易使数据变得更差。

——Philip B. Stark

(沈崇麟译校)

## 参考文献

Freedman, D. A., & Wachter, K. W. (2001). *On the likelihood of improving the accuracy of the census through statistical adjustment* (Tech. Rep. 612). Berkeley: University of California, Department of Statistics.

The 1980 U. S. Census [Special issue]. (1992). *Survey Methodology*, *18*.

The 1990 U. S. Census [Special issue]. (1993). *Journal of the American Statistical Association*, *88*.

The 1990 U. S. Census [Special issue]. (1994). *Statistical Science*, *9*.

The 2000 U. S. Census. (2001). *Society*, *39*, 2-53.

U.S. Census Bureau. (2001). *Report of the Executive Steering Committee for Accuracy and Coverage Evaluation Policy on Adjustment for Non-Redistricting Uses* (With supporting documentation, Reps.1-24). Washington, DC: Government Printing Office.

# 中心极限定理(Central Limit Theorem)

这条定理显示随着样本量 $n$ 的增加,即使总体值的分布是偏斜的,或其他非正态形的,统计值的抽样分布也将接近正态分布。

此外,随着 $n$ 的增大,抽样分布的伸展将变得狭窄,这就是说,正态分布的钟形将会变窄。它的使用与样本均值关系密切:样本均值的抽样分布【Sampling Distribution】接近正态分布【Normal Distribution】,集中在真的总体均值周围。由于较大的样本正态分布更加集中在总体均值的周围,因而将有较小的标准差。这一定理对统计推论至关重要,因为它意味着总体参数可用有各种置信度的样本统计值来估计。因为来自一个均值为 $\mu$ 的总体的重复样本的均值将形成一个围绕 $\mu$ 的正态分布,于是对于任何一个给定的样本,均值落入样本均值的某一范围内的概率是可以估计的。它并不需要实际上真地去进行

重复抽样。重复抽样产生正态和(较大 $n$)狭窄的,集中在总体均值周围的样本平均值分布这一知识,使我们得以在各种置信水平下,预测我们的样本均值接近真均值的程度究竟怎样。对于大样本而言,即使总体分布非常偏斜或离散,这个关系仍然存在。这一点非常重要,因为在社会科学中,一般许多有待研究的变量的分布(如收入)都是偏斜的。大的 $n$ 可能意味着估计均值至少要有 30 个个案,但对其他的统计量则要 500 个以上的个案。

为了阐释这一定理,图 1 显示了一个偏斜分布的例子。这是一个截删的英国收入分布。数据来自 1990 年代的英国全国调查。从这一分布我们取了样本量为 $n = 10$ 的 60 个左右的随机样本,得到的分布显示在图 2 中。这一分布比图 1 的偏斜程度要小,但似乎并不是完全正态的。

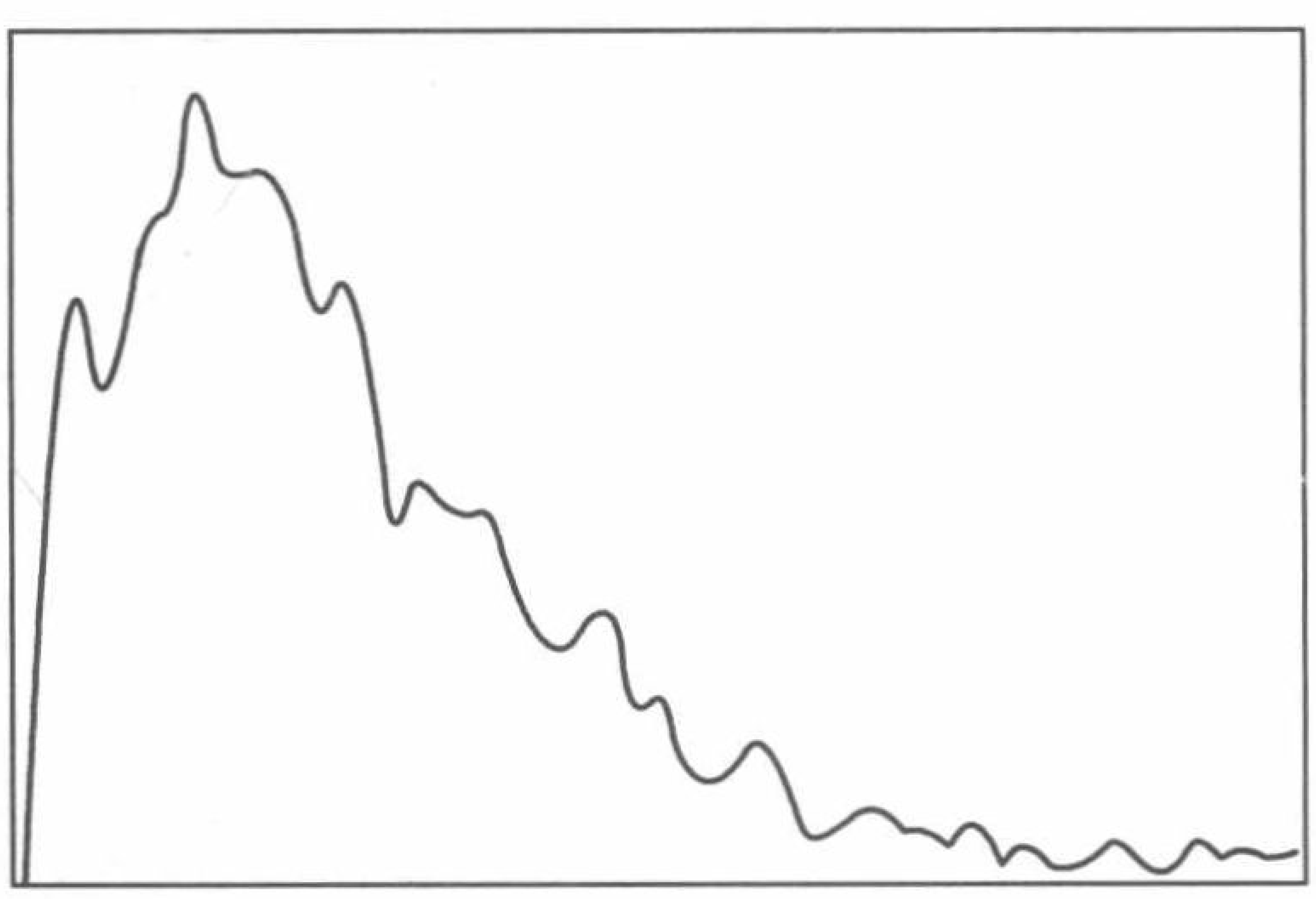


图 1 偏斜分布图例

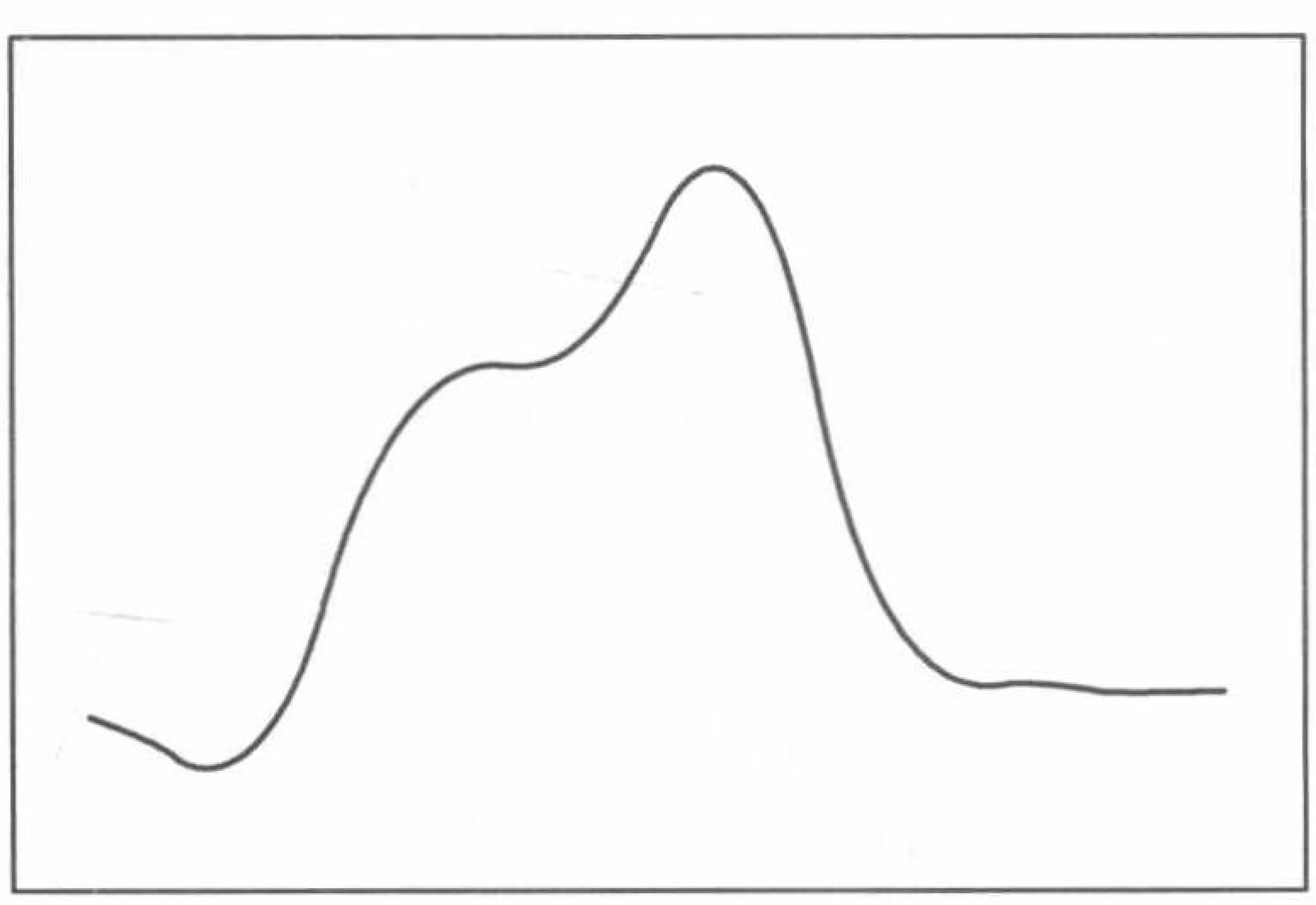


图 2 来自图 1 显示的分布的样本均值的分布,样本量 = 10

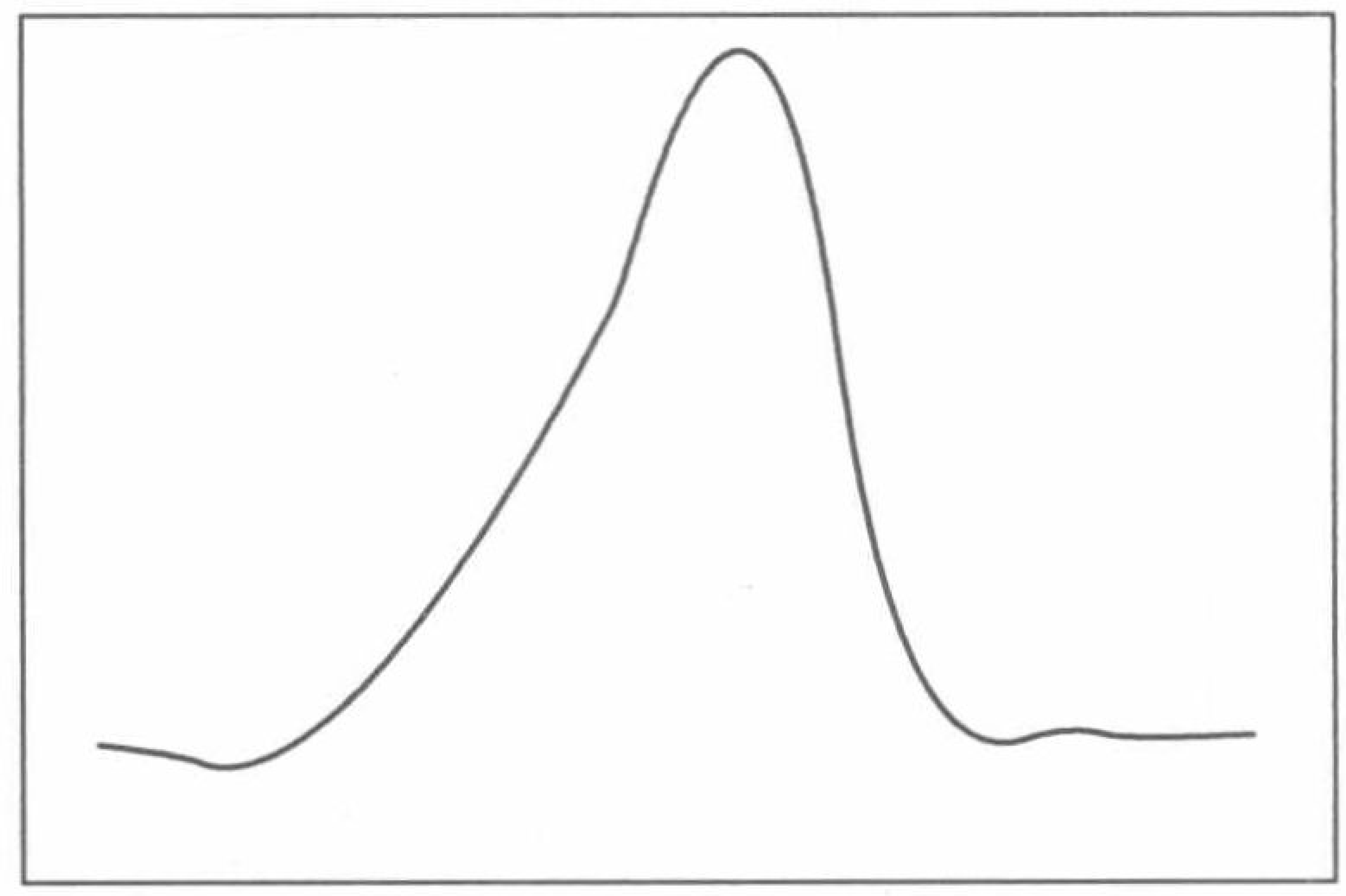


图 3 来自图 1 显示的分布的样本均值的分布,样本量 = 100

如果再进一步从图 1 显示的分布中取 60 个重复样本,这次的样本量为 100。这些样本均值形成的分布清楚地显示在图 3 中,不难看到它已经开始接近正态分布。

## 例子

由于有了这一定理,我们就可以估计总体的参数,例如总体均值,譬如估计英国女性非全职雇员的薪金。1980 年代后期,一个来自一个特定研究总体的,样本量为 179 个这样一种女性的随机样本,给出了每小时薪金 2.57 英镑的样本均值和 1.24 英镑的样本标准差(SD)。我们可以用样本标准差估计总体的标准差,它给出的标准误差为 $1.24/\sqrt{179} = 0.093$。根据计算得到的正态分布的概率,我们得知总体均值落在 $2.57 \pm (1 \times 0.093)$ 这一范围,即 2.48~2.66 英镑的概率为 68%;落在 $2.57 \pm (1.96 \times 0.093)$ 这一范围,即 2.39~2.75 英镑的概率为 95%;而落在 $2.57 \pm (2.58 \times 0.093)$,即 2.33~2.81 英镑的概率为 99%。

我们几乎可以肯定,样本均值覆盖的范围,即样本均值加上或减去三个标准误差覆盖的范围将包含总体均值的值。但问题很清楚,精确度和置信度之间需要有一个平衡。因此,若能有 95%的把握确信真均值落在了 2.39~2.75 英镑,我们就会感到比较满

意了，尽管我们知道有5%的可能性，我们的样本均值与真均值之间并不存在这种关系。也就是说，我们的样本中可能有一个样本，它的均值落入了钟形分布的两个尾部中的一个。

## 历史发展

中心极限定理可追溯到1880年，数学家皮埃尔·拉普拉斯给出的证明，尽管它本身是亚伯拉罕·棣莫佛的极限定理的一个发展。但是定理一经问世立即就产生了巨大的吸引力，并使我们对统计的理解向前大大迈进了一步。二项分布和泊松分布便是定理应用的两个最好的例子。然而，以奥古斯丁·柯西（Augustin Cauchy）的名字命名的，但也为泊松所确认的柯西密度却对我们提出了警告。这是一个分布的实例：对于有未定义的均值和标准差的这样一种分布而言，中心极限定理是不成立的。这种分布的存在表明：大样本法则存在例外，中心极限定理所基于的有关潜在的分布的某种假定也并非在所有的情况下都能得到满足。

然而，中心极限定理仍然是统计推论和估计的关键，它使我们能用大样本，根据正态性假设进行推论。可以证明，在那些与应用于单变量情形相同的条件下，中心极限定理同样也适用于多变量。

——Lucinda Platt

（沈崇麟译校）

＊也可参见钟形曲线【Bell-Shaped Curve】、正态分布【Normal Distribution】和抽样分布【Sampling Distribution】。

## 参考文献

Stigler, S. M. (1999). *Statistics on the table: The history of statistical concepts and methods*. Cambridge, MA: Harvard University Press.

Stuart, A., & Ord, J. K. (1987). *Kendall's advanced theory of statistics: Vol. 1. Distribution theory* (5th ed.). London: Griffin.

# 集中趋势（Central Tendency）

参见集中趋势量度【Measures of Central Tendency】。

# 形心法（Centroid Method）

形心法是一种凝聚聚类方法（agglomerative clustering method），该法根据在聚类时使用的变量上的类的形心（即多维度方法）对类别中的相同点（或相异点）进行定义。尤其需要指出的是，形心法是一种系列凝聚层次的聚类，或SAHN算法【Algorithm】（Sneath & Sokal, 1973），也即一种所有的要被聚类的实体最初都是分离的，然后一步一步地合并，形成高度有序的类的方法。形心法最初是由索卡和米切纳（Sokal & Michener, 1958）提出的。

形心法输入的是一组$N$个要聚类的实体或"个案"，这些个案中的每一个（表示为$O_i$）都是一个在一组相关的数字描述的变量$X_{i,1}, X_{i,2}, \cdots, X_{i,M}$上的多变量的观察。换言之，

$$O_i = [X_{i,1}, X_{i,2}, \cdots, X_{i,M}]$$

形心法的输出是将个案分配给一组分层(和嵌套)的聚类。一般将这种分层的聚类的解用一张树图【Tree Diagram】或系统树图(dendrogram)表示。

## 算法述要

与任何一种凝聚聚类法一样,形心法的算法是一步一步进行迭代的。在第一步,共有 $N$ 个实体或个案要进行聚类。然后,在每一步都有两个同类合并,产生一个聚类;最后,这个类本身便被看作一个新的个案(而构成它的两个个案则被去掉)。因此,算法的第二步将在 $N-1$ 个个案的缩减组上进行。

在第一步,算法开始按两个个案之间的欧几里得距离(或欧几里得距离的平方)定义每对个案 $O_i$ 和 $O_j$ 之间的相异点:

$$d_{i,j} = \left[ \sum_{m=1}^{M} (X_{i,m} - X_{j,m})^2 \right]^{\frac{1}{2}}$$

注意,个案 $O_i$ 在变量 $X_1, \cdots, X_M$ 上的值,可看作有一个成员的微小的类的形心。这些距离的计算结果是每对个案之间的相异点矩阵 **D**。第一步顺序凝聚算法像下面这样连续进行。有着最小相异点的成对个案 $O_i, O_j$ 被选来进行合并,而这两个个案被合并成一个聚类,它可看作一个新的类 $O_{N'}$。这个新个案 $O_{N'}$ 将加入那组要进行聚类的个案,而它的构成个案 $O_i$ 和 $O_j$ 则要被去掉。换言之,多变量观察 $O_i$ 和 $O_j$ 被单个的新个案 $O_{N'}$ 取而代之了,它们的数据值就是类 $O_{N'}$ 的形心:

$$O_{N'} = [\overline{X}_{N',1}, \overline{X}_{N',2}, \cdots, \overline{X}_{N',M}]$$

这时,迭代算法的第一步便完成了,而下一步则开始了。注意现在的个案数是 $N-1$,而非 $N$(因为两个个案已被一个单独的类 $O_{N'}$ 取代了)。对于这种缩减的邻近矩阵(proximity matrix),新类 $O_{N'}$ 到每一其他个案 $O_k$ 的相异点,被定义为从这一聚类的形心到个案 $O_k$ 在变量上的值的向量之间的欧几里得距离。也就是

$$d_{N',k} = \left[ \sum_{m=1}^{M} (\overline{X}_{N',m} - X_{k,m})^2 \right]^{\frac{1}{2}}$$

第二步继续进行,选择新的最近的个案对(即一对有最小相异点的个案),将它们合并。在第二阶段,以及随后的各个阶段,两个要被合并的个案中的任何一个都有可能是一个类。两个最近的个案被合并了,这个组的个案数就又减少了一个,并用这个新的聚类的形心再来计算邻近矩阵。

这些按顺序进行的步骤被不断重复,直至剩下两个个案为止。这两个个案可能被合并成一个包含所有个案的单个的类。

### 加权和未加权形心法

上面介绍的算法忽略了一个复杂问题。这个问题是,在被合并的两个个案包括不同数目的原 $N$ 个个案时,形心应该如何计算?在最为普遍的算法版本中,形心是用来自组成那个类的原始个案组的数据(等加权)计算的。斯尼思和索卡(Sneath & Sokal, 1973)把这一版本叫作未加权对组形心法(unweighted pair-group centroid method)。另一方面,如果一个聚类的形心是按两个要合并的个案的简单多维度的平均数计算的,不考虑组成这两个当前聚类或个案的原个案数,那么较大的聚类中那些加权的个别个案的效应将小于较小聚类中的个案。算法的这一变换叫作加权对组形心法(weighted pair-group centroid method)。

## 一个例子

表 1 是一个数值的例子。在这一假设应用中,数据由 5 名学生的数学和语文考试成绩组成,而在第一步则使用了欧几里得距

离量度，最小的距离（= 82.46）是在个案 $s03$ 和 $s04$ 之间。这两个个案被平均形成聚类 $c34$。在第二步，实体间最小的距离是在个案 $s01$ 和类 $c34$ 之间，于是这两个实体合并（用未加权法）在第三步形成类 $c134$。算法以这种方式进行，直至所有的个案被合并到一个单独的类 $c^*$。

表 1

| 步　骤 | 主　体 | 数　学 | 语　文 | 最小距离 |
|---|---|---|---|---|
| 1 | $s01$ | 600 | 380 | $d(s03, s04) = 82.46$ |
| | $s02$ | 800 | 790 | |
| | $s03$ | 760 | 330 | |
| | $s04$ | 740 | 410 | |
| | $s05$ | 520 | 780 | |
| 2 | $s01$ | 600 | 380 | $d(s01, c34) = 150.33$ |
| | $s02$ | 800 | 790 | |
| | $c34$ | 750 | 370 | |
| | $s05$ | 520 | 780 | |
| 3 | $c134$ | 700 | 373.33 | $d(s02, s05) = 280.18$ |
| | $s02$ | 800 | 790 | |
| | $s05$ | 520 | 780 | |
| 4 | $c134$ | 700 | 373.33 | $d(c134, c25) = 413.61$ |
| | $c25$ | 660 | 785 | |
| 5 | $c^*$ | 684 | 538 | |

### 非单调

单调（monotonicity）是许多顺序凝聚层次聚类法，如单链或全链聚类，都要满足的一种性质。为了定义单调，令 $O_{N'}$ 是一个由个案 $O_i$ 和 $O_j$ 合并而成的一个聚类。如果新类 $O_{N'}$ 的相异性 $d_{N',k}$：对任何其他的个案 $O_k$，总是等于或大于另两个相异性 $d_{N',i}$ 和 $d_{N',j}$，那么层次聚类法就满足了单调。形心法无论是加权还是未加权的，都未能满足单调（Anderberg，1973；Milligan，1979）。这种方法因为未能满足这一性质而受到批评，因为只有满足了单调，我们才能确保用来表示分层聚类结构的树形图，可以用表示在算法的一个特定水平或阶段的聚类之间的相异性的每个节点的高度来绘制。

### 在社会科学中的适用性

因为形心法要求输入一组有多个假定都是数值型的变量的观察，所以在数据是由病人记录或有几个子尺度（subscales）的考试成绩量度的主体的样本构成时，形心法都

是适用的。用欧几里得距离或顺序欧几里得距离定义聚类之间的相异性,意味着变量或子尺度应该都是以可比尺度量度的。在社会科学中,形心法的使用较其他的 SAHN 技术,如 Ward 法要少。

——James E. Corter

(沈崇麟译校)

## 参考文献


Anderberg, M. R. (1973). *Cluster analysis for applications*. New York: Academic Press.

Milligan, G. W. (1979). Ultrametric hierarchical clustering algorithms. *Psychometrika*, *44* (3), 343-346.

Sneath, P. H. A., & Sokal, R. R. (1973). *Numerical taxonomy: The principles and practice of numerical classification*. San Francisco: Freeman.

Sokal, R. R., & Michener, C. D. (1958). A statistical method for evaluating systematic relationships. *University of Kansas Science Bulletin*, *38*, 1409-1438.


# 其他条件不变(Ceteris Paribus)

Ceteris paribus 是一句拉丁语短语,意思是"在其他条件相同的情况下"。因此,它指的是在确定因果关系,或一个变量对另一个变量的效应时,将相似物与相似物进行比较的过程,故可被看作一种其他条件不变。例如,我们可能对某种形式的培训是否对以后工资的增加有作用这一问题感兴趣。而为了确定这一特定的培训效应,我们需要控制一些个体的特征,如性别、教育程度、工作史、年龄等;此外,也许还有当地的劳动力市场条件。然后我们才能在那些其他因素上相配,即有着类似的教育程度、工作史等,而只是在受到其他额外的培训上有所不同的个体身上,确定培训对个体的效应。如果在这些条件得到控制时,培训仍然有显著的净效应,我们才能说在其他条件相同的情况下,培训提高了工资。实际上我们在用统计方法进行相似物与相似物的比较时,为了能够确定效应,并不需要我们的样本数据中的个体完全匹配,其他条件完全相同。相反,多变量法估计的是每个不同因素彼此的净效应。于是培训的效应将在所有其他变量的某一固定的位置上进行估计(如在某一特定教育水平上男人与男人比较等),即使这些固定位置的某些组合在我们的数据中并未发生。

例如,我们可能对女性的工资是否低于男性这一问题感兴趣。表 1 显示来自一个线性回归【Linear Regression】的结果。回归使用的数据来自 1980 年代后期,英国钟点工作工资的调查。我们要回答的问题是,平均而言女性每小时的工资是否低于男性?由表 1 的模型 1 可见,答案似乎是肯定的。她们似乎每小时比男性少赚近 2 英镑。

**表 1 回归分析显示的钟点薪酬的平均性别效应**

| | 模型 1 | 模型 2 |
|---|---|---|
| 常量 | 4.97 | 3.94 |
| 性别<br>(参照类=男性) | | |
| 女性 | −1.96*** | −1.79*** |
| 职业<br>(参照类=体力劳动) | | |
| 专业人士 | | 2.23*** |
| 中层人士 | | 0.78** |

续表

| | 模型 1 | 模型 2 |
|---|---|---|
| 教育程度（参照类=未受过正规教育） | | |
| 职业学校和中学 | | 0.16(*ns*) |
| 大专及以上 | | 0.93*** |

注：** $P<0.01$；*** $P<0.001$；*ns*=在 0.05 水平统计上不显著。

然而，有人可能会说女性之所以比男性赚得少，是因为她们所处的职业地位与男性不同，以及她们的平均受教育程度偏低，而非关键因素性别本身的作用。因此，为了进行类似物与类似物的比较，我们对职业水平和教育水平加以控制，得到的结果列在了表1的模型2中。我们现在看到，在其他条件相同的情形下，性别的效应减少到了每小时差1.79英镑。另一方面，我们看到从事专业工作的男性和女性，比从事体力劳动的每小时多赚2.23英镑。因此，每小时薪酬的性别差异，虽然有，但却只有一部分可归结于性别中存在的职业和教育模式差别。在其他条件相同的情形下，我们仍然可以说，女性的确比对应的男性赚得少，因而这个差异在统计上是显著的。

——Lucinda Platt

（沈崇麟译校）

＊也可参见控制【Control】、多变量分析【Multivariate Analysis】。

## 参考文献

Lewis-Beck, M. S. (Ed.). (1993). *Regression analysis*. London: Sage/Toppan.

# CHAID

卡方自交互探测(Chi-Squared Automatic Interaction Detection，CHAID)是一种基于树形结构的分割技术，在涉及为数众多的分类解释变量的分类因变量分析中很有用。它的基本算法是将样本(最初的“母”群)连续拆分成越来越小的，在因变量方面显著不同的子群，并因此而长出一棵树。这一过程一直继续到诸停止规则中的一条被触发为止。

在任何给定的分析点，为了能将母群拆分成两个或更多个子群，要根据那个被确定与因变量关系最为显著的($p$ 值最小)预测因子对母群进行评估。在原算法中(Kass，1980)，预测因子的 $p$ 值，是用将那些和因变量关系无显著不同的类别合并后的卡方独立性检验计算的，而调整则是用恰当的 Bonferroni 乘数。

预测因子被用户分为“自由的”“单调的”或“浮动的”，以创建那些适合合并的分类；合并算法将分类合并，以逐步的方式，每次合并两个。一个自由类的预测因子可以自由地和那个预测因子的其他任何类合并。对于单调变量，只有相邻的类别才可以合并。浮动变量也作为单调处理，除了最后一类经常用来表示丢失数据之外，它可以和任何其他的类合并。

它的扩展包括以下几种：

1.在 $Y$ 为有序的类别时，独立性检验被无 $Y$ 关联的检验所代替(Magidson，1993)。这一检验评定均值 $Y$ 得分之间差异的显著性类似 ANOVA，不过没有正态性假定。

2.对于连续的因变量，用 $F$ 统计量取代卡方统计量。

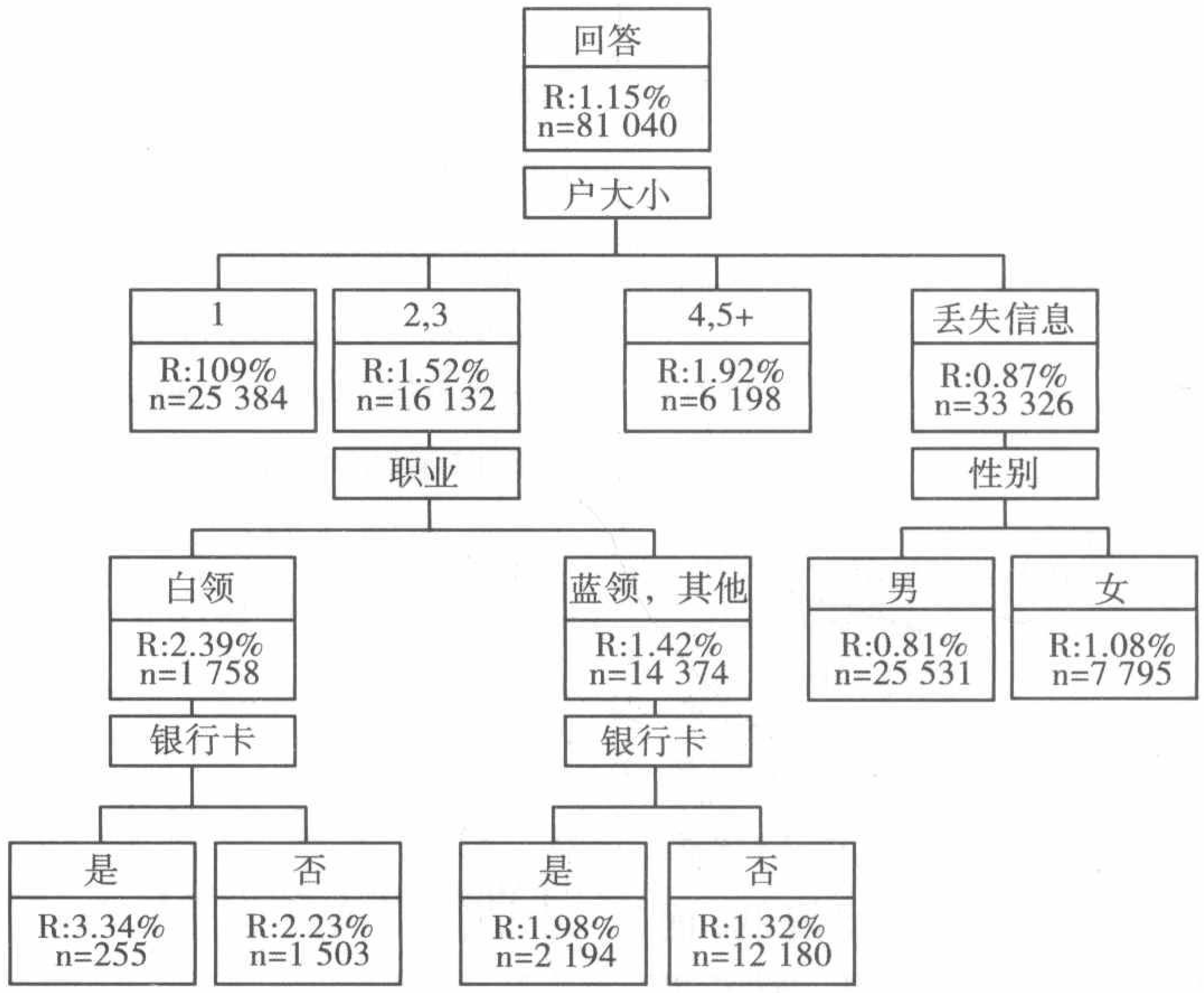



**图 1　卡方自交互探测**

3.合并算法被搜遍所有可能性的穷举搜索(exhaustive search)所取代(Biggs, De Ville, & Suen, 1991)。

4.用一个混合的 CHAID 加强模型的效力,如 logit 或有序 logit,形成树的结构或潜类,以将这种分析扩展到多因变量。

在回应是多量度(赢利能力、不同目录项的购买、购买数量)的,或剖视从一个潜类分析【Latent Class Analysis】得到结果片段时,我们将会使用潜类/CHAID 的混合分析。首先,进行 LC 分析,以得到潜在的潜变量,并将它递交给 CHAID 作为因变量,然后再把它作为人口或其他的外生变量的函数进行分析。

结果树上的每个节点预测每个原因变量之外的潜类片段,后者用 LC 模型的参数求得。这种类型的分析也会导致适合于所有的原因变量的预测因子类别的常拆分(common splitting)。

举一个 CHAID/logit 混合的例子,我们对来自马吉德松(Magidson,1993)的数据进行再分析。使用原算法的最好片段是户大小=2,或职业="白领",达到 2.39%的回答率。图 1 显示来自混合的 CHAID 的结果,其中 logit 建模增加的效力使我们得以将 CARD 的主效应与两个职业节点拟合,得到一个有 3.34%的预测回答率的额外片段。

——Jay Magidson

(沈崇麟译校)

## 参考文献

Biggs, D., De Ville, B., & Suen, E. (1991). A method of choosing multiway partitions for classification and decision trees. *Journal of Applied Statistics*, *18*,49-62.

Kass, G. (1980). An exploratory technique for investigating large quantities of categorical data. *Applied Statistics*, *29*(2), 119-127.

Magidson, J. (1993). The CHAID approach to segmentation modeling: CHi-squared Automatic Interaction Detection. In R. Bagozzi (Ed.), *Handbook of marketing research* (pp. 118-159). London: Blackwell.

## 变化分(Change Scores)

变化分是在时点($Y_t$)量度的变量值,与在前一个时点($Y_{t-1}$)用同一单位量度的变量值之间的差。变化分也叫作差异分数【Difference Scores】,或用符号表示的 $\Delta Y$,这样

$$\Delta Y = Y_t - Y_{t-1} \quad (1)$$

变化分是定群数据【Panel Data】,或在同一个体身上收集一次以上数据分析的关键。大多数此类分析的目的是,建立与时间上的自变量水平或变化有关的 $Y$ 是怎样变化的模型。变化分也是试测【Pretest】/后测实验设计【Experimental Designs】中的主因变量【Dependent Variabl】。

一种使用比较普遍的分析变化分的方法是将 $\Delta Y$ 作为 $X_1$,$X_2$ 等自变量,以及 $Y$ 自身在较早的时间 $t-1$ 的值的函数建模,如下式所示:

$$\Delta Y = \beta_0 + \beta_1 X_t + \beta_2 Y_{t-1} + \varepsilon \quad (2)$$

式中的 $\beta$ 表示回归系数【Regression Coefficients】,而 $\varepsilon$ 则是方程的误差【Error】项。$X_{t-1}$和 $X$ 其他的滞后值效应也可以包括在模型【Model】中。$Y_{t-1}$ 可以加到变化分方程(2)的两侧,这将产生一个预测 $Y$ 在时间 $t$ 时的水平等价模型:

$$Y_t = \beta_0 + \beta_1 X_t + (\beta_2 + 1) Y_{t-1} + \varepsilon \quad (3)$$

$Y_{t-1}$或"滞后【Lagged】内生变量"通常都因各种原因而被包括在变化分模型中。但最重要的也许是纵贯【Longitudinal】数据中那种叫作趋均值回归【Regression Toward The Mean】的现象所致。回归效应反映的是某一时点上大的变量值的趋势,它与较大的随后的负向变化有关联,因为大的初始值一部分是由大的随机误差【Random Error】所产生的,而在下一个时期,它们可能比较小。如果我们可以证明前面的 $Y$ 值是后面的值的因果【Cause】,那么我们就能有实质性的理由证明将 $Y_{t-1}$包括进来是正当的。

变化分模型的估计是比较复杂的,因为 $Y$ 中存在随机测量误差【Measurement Error】,在这种情况下,$Y$ 中观察到的变化,将不能精确地反映时间上的变化。纠正量度误差一般需要两期数据的更多 $Y$ 的指标(即其他那些量度相同的概念或像 $Y$ 一样的潜变量【Latent Variable】的变量),或 $Y$ 的单指标模型的更多期的量度(Kessler & Greenberg, 1981)。另一个可能的问题是存在自相关【Autocorrelated】的干扰,致使 $Y_t$ 和 $Y_{t-1}$的误差项相关,这与在纵贯数据中经常发生的问题一样。如果那样的话,$Y_{t-1}$将和误差项相关,而这将需要估计带有一个工具变量【Instrumental Variable】或有关方法的模型。

变化分在试图控制固定效应【Fixed Effect】的模型中也有着突出的作用。也就是说,虽然某个给定的个体或单位的特征尚未为研究者观察到,但那些特征在时间上是稳定的,如果 $Y_t$ 是由一系列 $X$、未观察到的稳定的变量 $Z$ 和一个误差项决定的,那么 $Z$ 的潜在的偏倚效应可以通过构建一个变化分模型来消除。那就是从 $Y_t$ 的方程减去 $Y_{t-1}$的方程(或在多期数据中,从单位的总均值减去每年每个单位的 $Y$),如果在动态固定效应模型中将 $Y_{t-1}$作为一个自变量,那么它将必定与误差项相关,这样为了纠正这种潜在的偏倚,使用工具变量和多期数据就是必需的(Hsiao, 1986)。

——Steven E. Finkel

(沈崇麟译校)

* 也可参见差异分数【Difference Scores】。

## 参考文献

Hsiao, C. (1986). *Analysis of panel data*. Cambridge, UK: Cambridge University Press.

Kessler, R., & Greenberg, D. (1981). *Linear panel analysis*. New York: Academic Press.

# 混沌理论(Chaos Theory)

混沌理论讨论的是一种非线性动力学【Nonlinear Dynamics】性状,是不规则和振荡的。它在数学上与一组非线性的确定性动态模型相遇,无论模型的运行持续多久,其性状的延时模式都是非重复的。人们发现,有些使用非线性差分方程(nonlinear difference equations)的离散时间模型会在仅使用一个方程的条件下,表现出混沌的动态性。但是在连续时间模型中,满足混沌的可能性一般需要至少三个独立变量,一般来说,它需要一个由三个微分方程组成的互依系统。连续时间模型的这个要求在特殊情况下也可以改变,如果系统还有一个强迫振荡子(forced oscillators)或者时滞,混沌甚至会出现在只有单个方程的连续时间模型中。

混沌只在非线性的情境下出现。在多维设置中,这意味着在也涉及若干变量时,在一个方程至少一个项必须是非线性的。因为大多数非线性模型(几乎所有实质上有趣的模型)都没有解析解,所以它们必须用许多数字密集的方法来探究,需要使用计算机。因为混沌的广延性和非线性并不能保证它就会出现,一般来说,模型中的混沌的性状是通过计算实验发现的。我们要在实验中找到引起模型显示混沌性的变量范围和参数的值。

尽管我们当前确定和为混沌过程建模的能力还在发展之中,但混沌过程在实际的物质甚至是社会系统中可能都是相当常见的。发现物质和社会系统中的真实的混沌过程常常是相当困难的,因为在这样的系统中几乎总是存在着随机的【Stochastic】噪声,所以要把真的随机【Random】性状和混沌性状区分开来也不是那么容易。尽管如此,各种旨在挑选这些过程和问题的数学工具还是不断地被开发出来。

混沌有三个基本特征,分别是:(a)不规则的周期性;(b)对初始条件的敏感性;(c)缺乏可预测性。在任意一个混沌设置中,这些特征相互影响,产生了高度复杂的非线性的变量轨迹。不规则的周期性指的是被混沌驱动的变量的振荡运动没有一个重复的模式。考虑到不规则的周期性,人们通常用傅里叶分析、绘图技术以及其他方法来建立个案,以识别一些混沌过程(Brown, 1995a)。

就离散设置中的混沌而言:其中被研究得最充分的模型是非线性模型,它是Logistic 映射(logistic map)的一般形式,梅(May, 1976)第一个考查了它的混沌特征。这种一般的 Logistic 映射是 $Y_{t+1}=aY_t(1-Y_t)$。在正确的条件下,这个映射可以产生标准的 S 形轨迹,它是所有 Logistic 结构的特征。但是,当参数 $a$ 的值足够大时,轨迹可能会发生振荡。例如,在将参数的 $a$ 值设为 2.8 时,模型的轨迹会在 $Y_{t+1}=Y_t=Y^*$ 的平衡值附近振荡并与此同时向平衡极限(equilibrium limit)收敛。但当我们将参数的 $a$ 值设定为另一个值,例如,设为 4.0 的时候,这时产生的纵向轨迹便不会平定于平衡极限,相反,将继续围绕平衡无规则地振荡,颇似一种由确定【Deterministic】过程引

起的随机方式。

最著名的体现了混沌性状的连续时间模型是所谓的“洛伦兹吸引子”(Lorenz attractor)(Lorenz, 1963)。这个模型是一个互依的非线性系统,涉及三个一阶微分方程。而它最初是被用来分析气象现象的。有时具有强制振荡子(forced oscillators)的系统是展现混沌或接近混沌的(即看似随机的)纵向特征的好的候选模型。在社会科学中,布朗(Brown,1995b, Chap. 6)开发并探究了这种模型。这个模型是一个由四个互依的微分方程组成的非线性系统,它对说明白宫(或者其他重要的政府机构)的交替党派控制的参数使用了一个强制振荡子。从对环境的纵向破坏、公众对环境破坏的关心,以及清理环境的成本等方面来考查这个系统的动态。有些参数值的变化产生了各种各样的稳定或不稳定的非线性动态性状,包括一些看似随机的性状,但实际却常常都与混沌有关联。

——Courtney Brown

(杨可译　沈崇麟校)

## 参考文献


Brown, C. (1995a). *Chaos and catastrophe theories.* Thousand Oaks, CA: Sage.

Brown, C. (1995b). *Serpents in the sand: Essays on the nonlinear nature of politics and human destiny.* Ann Arbor: University of Michigan Press.

Lorenz, E. N. (1963). Deterministic non-periodic flow. *Journal of Atmospheric Science*, *20*, 130-141.

May, R. M. (1976). Simple mathematical models with very complicated dynamics. *Nature*, *26*, 459-467.


# 卡方分布(Chi-Square Distribution)

卡方分布是一种由标准正态变量的平方和形成的分布。尽管正态分布变量的和有一个正态分布【Normal Distribution】(有一条钟形曲线【Bell-Shaped Curve】),但它们的平方和却没有。我们常用的卡方分布类型是中心卡方分布(central chisquare distribution)。我们下面的讨论将集中在这种类型的卡方分布上。因为正和负两种值的平方都是正数,所以曲线只包含非负数,曲线的左尾从零附近开始,曲线是向右偏斜的。曲线由它的自由度【Degrees of Freedom】(*df*)设定,所谓自由度是那些将要取平方和的未受约束的变量数。曲线的均值等于 *df*,标准差【Standard Deviation】等于两倍的 *df* 的平方根。曲线的最高点在 *df*-1 处。随着 *df* 的增加,分布的峰形也随之增高,偏斜也减少,但伸展却更广,因此标准差也随之加大,不过最远的左点仍然是 0。因此,一个有 5 个 *df* 的分布,其均值为 5,标准差为 3.1,而一个有 10 个 *df* 的分布,均值为 10,标准差为 4.5。在高 *df* 值处,曲线趋向于正态分布,在这种(罕见的)情况下,它有可能恢复正态概率,故而不适合再用那些为卡方分布计算的值。卡方分布右尾的概率,是右尾部分某一值之上的面积除以表示卡方值随机分布的可能性的曲线下的总面积。如果这个尾部的面积超过了给定的卡方值表示的面积,譬如说是曲线下的面积的 5%,这时我们就有 95%的把握确信被检验的关联是成立的。

## 应用

某些范围的卡方值和 *df* 的这些概率已经被计算出来并制成了表格。而那些普遍刊印的表格则提供了某个概率水平和某个

*df* 的卡方值。例如,诸如这样的表格告诉我们,在 4 个 *df* 卡方值为 9.49 时,我们可有 95%的信心。图 1 显示了 4 个 *df* 的卡方分布(均值为 4,标准差为 2.83)。右尾的概率在 5%处,该处作为示例的截线取值 9.49。这意味着,在有接近卡方分布的统计量时,我们可以将检验统计值和这样的表格进行比较,并确定它的显著水平。

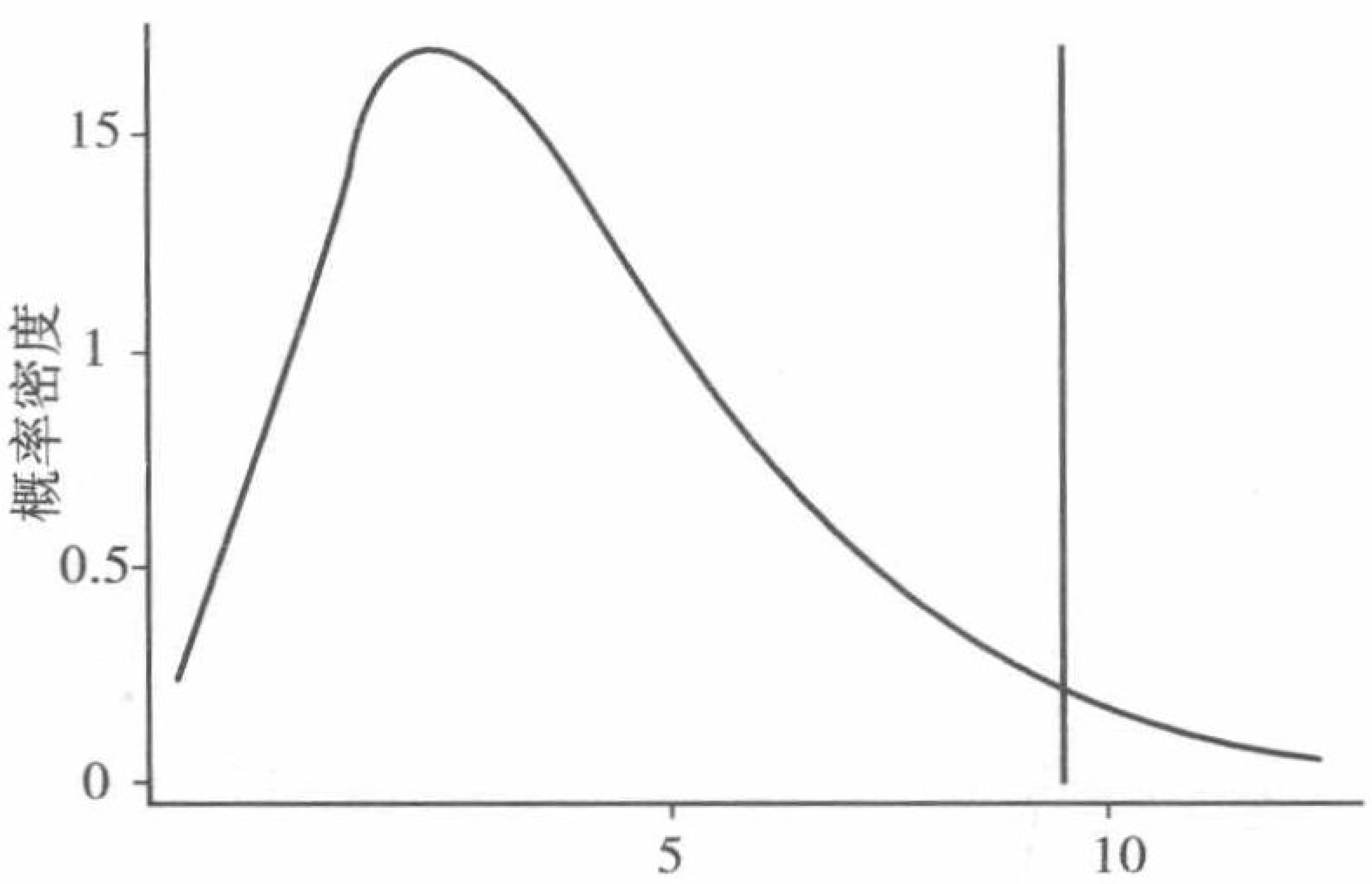



图 1　一个有 4 个 *df* 的卡方分布的例子

最常见的统计量来自皮尔逊卡方检验【Chi-Square Test】。而这样的检验并非只使用卡方分布。其他检验观察和"期望"或理论值之间的关系的统计量(如量度拟合优度的统计量)都接近卡方分布。例如,似然比检验统计量也称为 $G^2$ 或 $L^2$,则源自最大似然法,接近一种卡方分布。它通过基于虚无假设的最大似然(负的 2 倍的对数似然)和基于得到的实际数据值的最大似然之间的比的(自然)对数变换求得。$G^2$ 是对数线性模型最常使用的统计量。在实际工作中,卡方和 $G^2$ 统计量虽然有区别,但常常给出类似的结果,并得出有关给定模型的拟合优度类似的结论,不过对于小样本而言卡方则更稳定。

## 例子

例如,表 1 列出了英国 1980 年代后期两年的分阶级选举数据。虚无假设是格频数,是随机分布的。而 4 个 *df* 的卡方和 $G^2$(或最大似然比卡方)分别为 84.7 和 88.4。很清楚,这些值对于基于 4 个 *df* 的卡方而言是很高的值,因而表明这个模型完全不可能反映现实。我们对模型加以约束,以使投票和阶级的期望值可随着彼此的变化而变化,但跨两年的变差变成了常数,这样产生的模型,有 3 个 *df*,皮尔逊和似然比的值都为2.1。很清楚,诸如这样一个值,在和有 3 个 *df* 的卡方分布相比较时是很可能的($p$ = 0.54),因而它表示有相对简约的优度拟合。卡方和 $G^2$ 的值(约为 1 个 *df* 的 86 和 82)清楚地表示有非常显著的变化 。

表 1　英国 1980 年代后期两年的分阶段选举数据

| 年 | 阶　级 | 投　票 | | 合　计 |
|---|---|---|---|---|
| | | 工　党 | 保守党 | |
| 第 1 年 | 工人阶级 | 50 | 20 | 70 |
| | 中产阶级 | 10 | 40 | 50 |
| 第 2 年 | 工人阶级 | 70 | 30 | 100 |
| | 中产阶级 | 20 | 80 | 100 |
| 合　计 | | 150 | 170 | 320 |

## 历史

若干卡方分布的概率是由帕林・埃尔德顿于 1901 年制成表格的。卡尔・皮尔逊以及他的同事们在 20 世纪初,使用这些表格展示了他们的卡方检验。而现在人们普遍认为皮尔逊对 *df* 数计算的设定是错误的,尽管费希尔曾试图予以纠正,但是被拒

绝了，只是在后来它才被承认是正确的解决方案。所以说皮尔逊在评估分布的概率值时，用的 *df* 太多。最大似然比的源头可上溯到 17 世纪，虽然一直到 20 世纪后半叶，由于诸如里奥·戈德曼和斯蒂芬·苏伯格等研究者的工作，logit 和 log-linear 模型，以及与之相伴的拟合优度【Goodness-of-Fit】统计量得到发展和使用，才使最大似然比得到真正使用和发展。

——Lucinda Platt

（沈崇麟译校）

* 也可参见卡方检验【Chi-Square Test】、自由度【Degrees of Freedom】、对数线性模型【Log-Linear Model】、概率比对数模型【Logit Model】。

## 参考文献


Agresti, A.(1996). *An inrtoduction to categorical data analysis.* New York: John Wiley.

Fienberg, S. E. (1997). *The analysis of cross-classified categorical data*. Cambridge: MTT Press.

Yule, G. U., & Kendall, M. G. (1950). *An introduction to the theory of statistics* (14th ed.). London: Griffin.


# 卡方检验(Chi-Square Test)

卡方检验是社会科学中最常用于类别变量显著性检验的检验方法。它是卡尔·皮尔逊在 1900 年前后作为一种用于评估在一张列联表【Contingency Table】中交叉制表的两个类别变量之间的相互关系的工具。检验将表中的实际值与在独立性条件下（即如果两个变量之间没有关系）的期望值相比较。每格的期望值是这样计算的，即将该格的行比例乘以列比例再乘以个案总数。例如，表 1 显示了假设的投票者的例子的期望计数。它是基于行和列的总数（叫作边缘分布）及总个案数得到的。它还显示观察的计数与由独立性决定的那些计数是不同的。在独立性情形中，投票行为并不会随性别变化。

**表 1　独立性条件下的不同性别的投票行为期望值**

| | 党派 A | 党派 B | 其他党派 | 总　计 |
|---|---|---|---|---|
| 男　性 | 0.46 · 0.5 · 70 100 | 0.46 · 0.3 · 70 100 | 0.46 · 0.2 · 70 100 | |
| 期　望 | 16 123 | 9 674 | 6 449 | 32 246(46%) |
| 观　察 | 13 669 | 13 811 | 4 766 | |
| 女　性 | 0.54 · 0.5 · 70 100 | 0.54 · 0.3 · 70 100 | 0.54 · 0.2 · 70 100 | |
| 期　望 | 18 927 | 11 356 | 7 571 | 37 854(54%) |
| 观　察 | 21 381 | 7 219 | 9 254 | |
| 总　计 | 35 050(50%) | 21 030(30%) | 14 020(20%) | 70 100(100%) |

表 2 不同阶级的收支平衡情况（英国 1980 年代）

| | 容易入不敷出 | 收支平衡 | 收大于支 | 合 计 |
|---|---|---|---|---|
| 专业人士/管理人员 | 124(84) | 76(86) | 57(87) | 257<br>26% |
| 中间阶级 | 73(75) | 84(77) | 73(73) | 230<br>23.3% |
| 工人阶级 | 125(163) | 171(167) | 204(169) | 500<br>50.7% |
| 合 计 | 322<br>32.6% | 331<br>33.5% | 334<br>33.8% | 987<br>100% |

当然，即使在总体中是独立的，样本中的观察值也不太可能完完全全地模拟期望值。卡方检验确定的是，以样本为根据的表格格子中的观察值和期望值之间的任何差异是否意味着总体中的真差异（即格计数是不独立的）。这一点是通过比较实际的格计数和如果跨解释变量的每一类别的比例是一致的期望的格计数做到的。卡方统计值是按

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e}$$

计算的。式中的 $f_o$ 为观察到的格计数，$f_e$ 为期望的格计数。然后，检验要将样本的容量和表格中的自由度考虑进去，以确定样本中的差异究竟是偶然所致，还是总体概率的反映（即差异是显著的）。自由度按（行数-1）×（列数-1）计算。卡方检验统计量的值可以与这些自由度的卡方分布【Chi-Square Distribution】联系起来，以确定观察值离期望值的这一散度发生在那一分布内的概率，或它是否可能落在它之外。

举个例子，表 2 显示了分阶级的收支平衡的难度，样本由 987 个人组成，地点是英国，时间在 1980 年代。比较每格的观察计数和期望计数（在括弧中给出）得到的卡方统计值是 47.9，自由度为 4。这一数字在 0.000 水平是显著的，因而说明财务困难和社会阶级是独立的概率小于 1‰。

卡方不提供有关关联强度或它的方向的信息，它只提供有关变量间的依赖性的信息。卡方值随样本量的增大而增大，因此一个大样本中低的 $p$ 值（高度显著），可能伴随的是变量间相当弱的关系。卡方检验也没有指出因变量的哪一个值随解释变量的变化而变化。有可能只有一或两个格的计数大大偏离它们的期望值。这一点可以通过更为细致地审视表格来确认。在表中的任何一个期望计数小于 5 的时候，不宜使用卡方检验。这种情况在小样本或表中含有大数目计数的格子时发生。在遇有这样的情况时，其他的备择检验，如费希尔的精确检验则更为有用。

——Lucinda Platt

（沈崇麟译校）

## 参考文献


Agresti, A. (1996). *An introduction to categorical data analysis*. New York: John Wiley.

Everitt, B. S. (1977). *The analysis of contingency tables*. London: John Wiley.

Fienberg. S. E. (1997). *The analysis of cross-classified categorical data*. Cambridge: MIT Press.

## 邹氏检验(Chow Test)

计量经济学家邹至庄(Chow,1961)在他的文章"线性回归模型回归系数组之间的相等检验"中,论证了广泛用于方差分析【Analysis of Variance】中检验一组均值的同质或相等性的 $F$ 检验和方差比检验【Variance Ratio Test】,可怎样被用于发现回归【Regression】模型的结构或参数的稳定性。

为了阐述这一检验,假设我们有1970—1995 年美国储蓄($Y$)和个人收入($X$)的以亿美元计的数据。

假设我们考虑下面的回归模型,或储蓄函数:

$$Y_t = \beta_1 + \beta_2 X_t + u_{it} \qquad (1)$$

式中的 $Y$ 是储蓄,$X$ 是收入,$u_{it}$ 是随机误差项,而 $t$ 则是时间指标。

我们可用普通最小二乘【Ordinary Least Squares,OLS】来估计前述储蓄回归的参数。邹氏检验提出的问题则是:回归系数 $\beta_1$ 和 $\beta_2$ 在整个样本期是否仍然是稳定的?在1981—1982 年,美国遭受了严重的经济衰退。经济衰退是否改变了人们的储蓄行为?为了回答这个问题,假设我们把样本数据分成了两个时期:1970—1981 年和 1982—1995年,并估计如下两个单独的回归:

$$\text{时期 1970—1981}: Y_t = \alpha_1 + \alpha_2 X_t + u_{2t} \qquad (2)$$

$$\text{时期 1982—1995}: Y_t = \lambda_1 + \lambda_2 X_t + u_{3t} \qquad (3)$$

式中的 $u$ 是特定时间的误差项。

如果储蓄与收入的关系在时间上是稳定的,那么我们可以期望 $\beta_1=\alpha_1=\lambda_1$(即截距在统计上是相同的)和 $\beta_2=\alpha_2=\lambda_2$(即斜率系数是相同的)。但是我们怎样来发现它们是相同的呢?邹氏检验采取以下的步骤估计上面给出的三个回归,并分别得到它们的残余平方和【Residual Sums of Squares,RSS】,$S_1(df=24)$,$S_2(df=10)$ 和 $S_3(df=12)$。注意,总共有 26 个观察,其中 12 个在 1970 —1981 年,14 个在1982—1995 年。

如果假设误差项 $u_2$ 和 $u_3$ 是正态分布的,有相同的方差[即 $\text{var}(u_2)=\text{var}(u_3)=\sigma^2$],并且独立分布,那么邹氏检验便按以下步骤进行:误差项独立的假设意味着两个样本(即样本时期)是独立的。因此加上 $S_2$ 和 $S_3$,并求得无约束的 $RSS_{ur}$,它有 $(n_1+n_2-2k)$ 个自由度。式中的 $n_1$ 和 $n_2$ 是在第一和第二时期的观察数,而 $k$ 则是模型中估计的参数数(本例中有两个)。

因为 RSS,$S_1$ 是在参数稳定的假设下求得的(即回归系数不随时期变化),我们可以把 $S_1$ 叫作约束残差平方和($RSS_r$),这一RSS 有 $(n-k)$ 个自由度,其中 $n=n_1+n_2$。现在如果参数是稳定的,那么 $RSS_{ur}$ 和 $RSS_r$ 就不应该有统计上的差异。但是如果参数没有稳定性,那么这两个 RSS 在统计上就会有差异。如果这一假设构成了邹氏检验,那么我们便可由下式来看到这点:

$$F = \frac{(RSS_r - RSS_{ur})/k}{(RSS_{ur})/(n_1+n_2-2k)} \sim F_{[k,(n_1+n_2-2k)]}$$

这就是说,$F$ 值可由如上所述的 $df$ 的 $F$ 分布求得。因此,如果应用中计算得到的 $F$ 值超过,譬如说,1%,5%或 10%的显著水平,我们就可以拒绝参数是稳定的假设。在目前的例子中,这就意味着储蓄函数随时间变化。另一方面,如果算得的 $F$ 值是不显著的,我们可以假设参数是稳定的。在这种情况下,我们便只需作一个合并了所有观察的回归就可以了。

邹氏检验的某些数字结果可以从美国

1970—1995年的储蓄和收入的数据中看到。该数据的测量单位为亿美元。前面讨论的各种残差平方和如下所列：

$$S_1 = 23\,248.30\ (df = 24)$$
$$S_2 = 1\,785.032\ (df = 10)$$
$$S_3 = 10\,005.22\ (df = 12)$$

因此在目前这个例子中，$RSS_r = S_1 = 23\,248.30$ 和 $RSS_{ur} = S_2 + S_3 = 11\,790.252$。将这两个值代入前面计算 $F$ 比率的公式，得到如下的结果：

$$F = \frac{(23\,248.30 - 11\,790.252)/2}{(11\,790.252)/22} = 10.69$$

在假设的条件下，这一 $F$ 值服从 $F$ 分布，分子和分布各有2个和22个自由度。1%的临界 $F$ 值（即犯第Ⅰ类错误的概率）是5.72。因为算得的 $F$ 值超过了这个临界值，所以我们可以说在样本时期内，美国的储蓄-收入关系曾经经历了一个结构性的变化：储蓄函数在两个子时期是不同的。

在使用邹氏检验的时候请记住，构成检验的假设必须满足，此外研究者还必须知道关系的变化发生在什么时候。在前面的例子中，如果我们不能确定储蓄-收入关系在1982年发生变化，那么我们就不能用邹氏检验。当然，我们可以选择不同的断点，并得到有关的 $F$ 值，从而由此选择能给出最大 $F$ 值的断点。对邹氏检验的这种修正叫作科万特似然比统计量（quandt likelihood ratio statistic）。

——Damodar N. Gujarati

（沈崇麟译校）

## 参考文献


Chow, G. (1961). Tests of equality between sets of regression coefficients in linear regression models. *Econometrica*, *28*(3), 591-605.

Gujarati, D. (2002). *Basic econometrics* (4th ed.). New York: McGraw-Hill.

Wooldridge, J. (2002). *Introductory econometrics* (2nd ed.). Cincinnati, OH: South-Western College Publishing.


# 古典推论（Classical Inference）

做一个推论就是通过某种可接受的推理形式，从一组前提得出结论。推论式推理是科学方法的基础。古典推论嵌在确定性框架中，并遵循演绎或归纳路数。也存在着其他形式的推论，但它们通常都嵌在概率性框架中。

在古典框架中，演绎推理得出的结论隐含在陈述的前提中，而归纳推理的论证则是从特殊到一般。在表述精当时，演绎论证可以显示其在逻辑上的有效或“真实”。另一方面，归纳推理并非逻辑的终结，因此归纳的结论不能无可争辩地显示出有效性。正因如此，有些逻辑学家拒绝把归纳推理作为正当的逻辑形式。

演绎推理的最好例证是古典三段论。它由一个大前提、一个小前提和结论组成。我们可以用符号将它写成如下的形式：

如果 $p$，那么 $q$

$p$
___

所以，$q$

举一个不十分贴切的例子：如果某某是一个人，那么某某终究是会死的；黛安娜是一个人，因此黛安娜终究是会死的。这样一种三段论式叫作演绎推理（modus ponens）。另一种在逻辑上可以接受的三段论是否否推论法（modus tolens），它以如下的方式进行推论，如果 $p$，那么 $q$；$\sim q$，所以，$\sim p$。因此，我们也许可陈述如下：如果某某是一个

人，那么某某终究是要死的；黛安娜是不会死的，所以黛安娜不是人。在这一个例子中，黛安娜可能是古希腊的女神。

有两种其他书写古典三段论的方式。但二者在推理时都会带来一些错误。例如，肯定后件的谬误则取如下的形式：如果 $p$，那么 $q$；~$q$，所以，~$p$。下面我们采用一个实例：如果某某是人，那么某某是会死的；黛安娜是会死的，所以黛安娜是人。这里的错误在于，会死并非是人的充分条件。实际上，黛安娜有可能是一只宠物猫。

第二种错误的结论，或否定前件的谬误取如下的形式：如果 $p$，那么 $q$；~$p$，所以，~$q$。我们再用一个实例来加以说明：如果某某是人，那么某某是会死的；黛安娜不是人，所以黛安娜是不会死的。再一次面临同样的情况，黛安娜可能是某人的一只猫，但猫虽然不是人，它们却与人同样有死亡的特点。后两种有问题的形式是人们在科学推理时常犯的错误。

正是否否推论法这一逻辑形式形成了科学假设的骨干。让我们假设 $H$ 隐含着某个结果 $O$。无论我们怎么经常观察到 $O$，都不能支持假设 $H$。实际上，假设是不能直接确证的，因为用 $O$ 来支持 $H$，就是一种肯定后件的谬误。然而，没有观察到 $O$，却为我们提供了否认 $H$，或拒绝假设的根据。因此，正如哲学家卡尔·波普尔所言，科学方法的标志性特点是推翻，而不是证明假设的能力。

虽然演绎方式形成了一个封闭的逻辑系统，但归纳却是一个开放的系统。归纳一般这样论证道：因为所有观察到的 $P$ 的实例都有特征 $C$，所以所有 $P$ 的实例都有特征 $C$。然而正如休谟十分清楚地指出的那样，所有观察的 $A$ 都是 $B$ 这一陈述，与某些未观察的 $A$ 不是 $B$ 这一陈述是不一致的。$A$ 并不是 $B$。但支持归纳法的人却争辩道，虽然问题不能在归纳上被证明，但是随着总体越来越多地被列举，归纳便取得了更多的证据。因此，人们越是接近所有观察的 $P$ 的实例都有特征 $C$，我们就会在所有 $P$ 的实例都有特征 $C$ 这一陈述中放入更多的证据。在 $P$ 是一个封闭的总体时，这种理由似乎是站得住的，但是在 $P$ 是开放或无限的时候，它就很难站得住了。

围绕逻辑推论存在的那些困难，并不一定意味着逻辑推论在科学中是没有用的。假设常常是通过推论过程——某种我们称为发现环节（context of discovery）建立的。在决定性框架之外通过由样本到总体的概括得出可靠的推论也是可能的。而这导出了现在我们称为推论统计学的艺术，在这门艺术中，在给出了某些约束之后，一个有效的推论的可能性是可以被我们估计的。

——Paul S. Maxim

（沈崇麟译校）

## 参考文献


Copi, I. M., & Cohen, C. (1990). *Introduction to logic* (8th ed.). New York: Macmillan.

Gensler, H. J. (2002). *Introduction to logic*. London: Routledge Kegan Paul.

Hempel, C. G. (1966). *Philosophy of natural science*. Englewood Cliffs, NJ: Prentice Hall.

Poundstone, W. (1988). *Labyrinths of reason: Paradoxes, puzzles and the frailty of knowledge*. New York: Anchor/Doubleday.


# 分类树（Classification Trees）

参见分类和回归树【Cart】。

# 临床研究(Clinical Research)

临床研究是医疗卫生领域旨在创造新知识、实验新的治疗方法的一种探索。就像所有形式的科学探索一样,临床研究也关注鉴定、描述、作出解释【Explanations】、解释的检验以及/或者干预/控制【Control】。这些关注点将依次和典型的方法关联。对前三者——探索性的考查而言,一般使用的是观察【Observations】、访谈【Interviews】和文献【Document】分析等质性研究方法,可以说是调查法的变种。解释检验尽管也是质性研究传统的一部分,但在实践中一般多用于流行病学的研究,并有同期群【Cohort】、横截面【Cross-Sectional】以及病例对照等设计。例如,一个叫作随机化控制试验【Randomized Control Trial, RCT】的实验【Experiment】设计就是检验某种疗法的疗效的(Miller & Crabtree, 1999)。医学卫生学科的文献,尤其是药学文献仍然主要由实验和流行病学的设计所主导。这种配置反过来也反映在了发展中国家对基础生物医学研究和应用生物医学研究拨款的先后次序。类似的模式也曾在有关临床实践的社会科学,例如心理学和社会学的发展之初起着主导作用。但是,现在这些学科以及其他社会科学中已经建立起了重要的质性数据流(qualitative and critical streams),新的调研方式也被用于临床问题。在"循证医学"(evidence-based medicine)时代,质性研究者提醒我们,治疗病人必须有多种依据——依据既可以来自随机化控制试验,也可以源自对病人生活、看法以及社会环境等细节的熟稔。

这种新的临床研究范式之一就是现象学【Phenomenology】,它关注的是理解和传递"生活世界"(life-world)中个人的经验和意图。与面向疾病的描述不同,现象学的报告呈现的是患病的经验。例如,佩卢西(Pelusi,1997)曾采访了乳腺癌的生存者,发现其共同的主题不仅是损失和不确定的未来,还有成长和启示。另一种范式是扎根理论【Grounded Theory】,它很适用于纵向的社会过程评估,其目的是在细致的观察、访问和文献分析的基础上产生理论。以洛伦茨(Lorencz,1991)对精神分裂症病人的研究为例,这些病人经历了住院治疗到社区的转变,即所谓"平常化(becoming ordinary)"的过程。人类学的民族志【Ethnography】基本传统也在临床工作中有广泛的应用。这不仅包括健康和疾病的文化决定论(Kleinman, Eisenberg, & Good, 1978),还包括医疗卫生机构的文化及其对临床服务表现的影响。一个著名的运用人类学的敏感例子是广泛的"医疗错误"问题,它使得人们越来越认识到,"责备和羞辱"的文化对于确认和理解医疗错误的根本原因是一个很大的障碍(Kohn, Corrigan, & Donaldson, 2000)。最后还有一个例子是参与行动研究【Participatory Action Research】,其旨趣是通过调研的"民主化"理解和促进社会变革。按这种模式,那些一般会被当作医疗研究的"被试"的人变成了共同的研究者。麦考利等人(Macaulay et al., 1997)对蒙特利尔附近的一个莫霍克族(Mohawk)社区的儿童进行研究,考察的重点是健康饮食和增加锻炼对于预防糖尿病的作用。在设计干预方法时,他们既采用了行为变迁模型,也用到了土著居民的学习方式的知识。

——Anton J. Kuzel

(杨可译　沈崇麟校)

## 参考文献

Kleinman, A. L., Eisenberg, L., & Good, B. J.

(1978). Culture, illness and care: Clinical lessons from anthropologic and cross-cultural research. *Annals of Internal Medicine*, *88*, 251-258.

Kohn, L. T., Corrigan, J. M., & Donaldson, M. S. (Eds.). (2000). *To err is human: Building a safer health system.* Washington, DC: National Academy Press.

Lorencz, B. (1991). Becoming ordinary: Leaving the psychiatric hospital. In J. M. Morse & J. Johnson (Eds.), *The illness experience: Dimensions of suffering* (pp. 140-200). Newbury Park, CA: Sage.

Macaulay, A. C., Paradis, G., Potvin, L., Cross, E. J., Saad-Haddad, C., McComber, A., et al. (1997). The Kahnawake Schools Diabetes Prevention Project: Intervention, evaluation, and baseline results of a diabetes primary prevention program with a native community in Canada. *Preventive Medicine*, *26*(6), 79-90.

Miller, W. L., & Crabtree, B. F. (1999). Clinical research: A multimethod typology and qualitative roadmap. In B. F.

Crabtree & W. L. Miller (Eds.), *Doing qualitative research* (2nd ed., pp. 3-32). Thousand Oaks, CA: Sage.

Morse, J. M., Swanson, J. M., & Kuzel, A. J. (Eds.). (2001). *The nature of qualitative evidence.* Thousand Oaks, CA: Sage.

Pelusi, J. (1997). The lived experience of surviving breast cancer. *Oncology Nursing Forum*, *24*(8), 1343-1353.

## 封闭式问题(Closed Question)

参见封闭式问题【Closed-Ended Questions】。

## 封闭式问题(Closed-Ended Questions)

调查的问题有两种:开放式问题【Open-Ended Question】由受访者自己回答,封闭式问题则给出了各种具体答案类别。尽管对两种问题相对而言的优缺点曾有过许多研究,但在任何调查中绝大多数的问题都是封闭式问题。

为了方便分析,所有涉及提问的调查研究都可以将被调查者进行分类,在封闭式问题中,研究者对答案的可能的恰当类别作了预判,并直接在提问的时候就提供给被调查者。这种形式存在许多潜在的问题,别的问题姑且不说,至少它一般都是以两分法来描绘这个"世界",因而有时候可能会对可能性的反映过于简单。这种限制在开放式问题里不存在,因此舒曼和普雷瑟(Schuma & Presser,1996)得出结论说,开放式问题常常能够提供更多的有价值的数据。不过,由于封闭式问题省去了对原话逐一代码【Code】的工作,从而大大降低了调查的费用,且可以更快捷地进入分析,因而成了人们优先选择的形式。

在设计或使用封闭式问题的时候,有许多标准化的建议,体现了对这一问题的大量研究成果。在询问调查对象是否同意一个单独的陈述时,采用强制性选择更可取。一般来说,除了测量强度,都应该提供一个中间的选项,而且也应该提供一个明确的"没有意见(弃权)"的选择。研究者应该采用多项选择来评估同样的题目,并对提问的顺序所产生的效应保持敏感。如果对提问的措辞或者答案选项可能产生的效应不太确定,那么研究的问卷中应该包括分割样本的

版本，以便进行对比分析。

——Michael W. Traugott

（杨可译　沈崇麟校）

## 参考文献

Converse, J. M., & Presser, S. (1986). *Survey questions*. Thousand Oaks, CA: Sage.

Fowler, F. J., Jr. (1995). *Improving survey questions: Design and evaluation*. Thousand Oaks, CA: Sage.

Schuman, H., & Presser, S. (1996). *Questions & answers in attitude surveys*. Thousand Oaks, CA: Sage.

# 聚类分析(Cluster Analysis)

社会科学数据【Data】集通常采用一组变量【Variable】的分析单位【Units of Analysis】上的观察形式。聚类分析的目的是根据某些变量的信息产生一个简单的单位分类，将它们归入不同的子组。这样的表述似乎有些含混不清，但却并非偶然，因为时至今日，聚类分析仍无一个正式的定义。尽管如此，对之作一个比较精确的定义还是可能的。聚类问题求解的任务是，将 $n$ 个单位聚类成 $r$ 个聚类（$r$ 远远小于 $n$），这样，一个聚类中的单位都十分相似，而不同的聚类的单位则是不相似的。换句话说，这些聚集起来的类有着彼此分开的同质的聚类。聚类分析这一标题包含着一组各种各样的解决聚类问题的工具（Everitt, Landau, & Leese, 2001）。在大多数时候，这些工具都用于数据的归纳【Inductive】性探索。我们希望聚类能使我们得以洞察数据的结构、单位的性质和产生变量的过程。例如，城市可依据社会经济和人口特征进行聚类。人可以根据他们的心理属性或拥有的其他属性聚类。

## 聚类分析的发展

在1980年以前，许多聚类问题是分不同的学科分别解决的，因此进步十分缓慢。在1960年代早期，人们开始探索聚类问题的通用解决方法，并在这方面取得了许多进展。索卡和斯尼思对聚类问题作了深入的讨论（Sokal & Sneath, 1963），并帮助建立了将聚类分析作为一种数据分析领域的框架。聚类问题的设定并不困难。无论是聚类问题表达的数学基础，还是大多数聚类问题的求解都不困难。它的困难在于聚类问题的求解过程中计算的复杂性。正因如此，这一领域的发展主要为计算技术的进步所驱动。一般来讲，聚类技术的发展使我们得益匪浅，因为有用的聚类使实质性的解释变得丰富。此外，许多技术的发展则源自在新的领域的大量探索性的实质应用。现在有许多国家的聚类分析师团体，都通过“国际分类团体联合会”（International Federation of Classification Societies）联系在一起。

## 求解聚类问题

一般来讲，聚类问题可以被阐述为建立一个（或多个）有着在所有可能的聚类上定义精当的判别函数最小值 $r$ 的 $r$ 类的聚类。判别函数提供了一个适合于所有聚类的量度。而在实践中，判别函数常常是一个隐式或被忽略的。在大多数应用中，聚类是一种分解，但“模糊聚类”完全有可能会有重叠的类。一旦分析单位已被选定，聚类分析大概要通过以下五个步骤来实施：

1.量度相关变量（可以包括定量变量【Quantitative Variable】和分类【Categorical】

变量,某种标准化形式可能是必需的);

2.为适当的(不)相似性量度建立一个(不)相似性矩阵【Matrix】;

3.通过聚类算法建立一个或多个聚类;

4.对得到的聚类进行评估;

5.对聚类进行有实际意义的解释。

虽然每一步都充满了危险,但第二步和第三步是最危险的,而第四步则常常被忽略。在第二步,不相似性量度[例如欧几里得(Euclidean)、曼哈坦(Manhattan)和明科夫斯基距离(Minkowsky distances)]或相似性量度(如相关【Correlation】和相配系数【Coefficient】)可能会被使用。量度是至关重要的:不同的量度可能会导向不同的聚类。在第三步,有许多可用于建立聚类的算法。每对(量度和算法)的选择,原则上可能会导向不同的单位聚类。

分层聚类可以取凝聚和区分两种形式。凝聚聚类开始于每一个单位在它自己的聚类中,然后系统地合并单位和类,直至形成一个单独的聚类。区分分层聚类进行的方向则与之相反。在这些分类中,有着多种选择。其中最为普遍的三种是:单连接(single-link)、平均连接(average-link)和全连接(complete-link)聚类。在单连接中,算法在计算组之间的(不)相似性时,是按两个组中最接近的单位的(不)相似性计算的;在全连接聚类中,算法使用的是两组中距离最远的单位;而在平均连接聚类中,算法使用的是两组单位的平均(不)相似性。沃德法(Ward's method)也是一种在合并两个类时使用普遍的方法。如何在这些方法中作出选择是至关重要的,具体的选择方法则如图 1 所示。

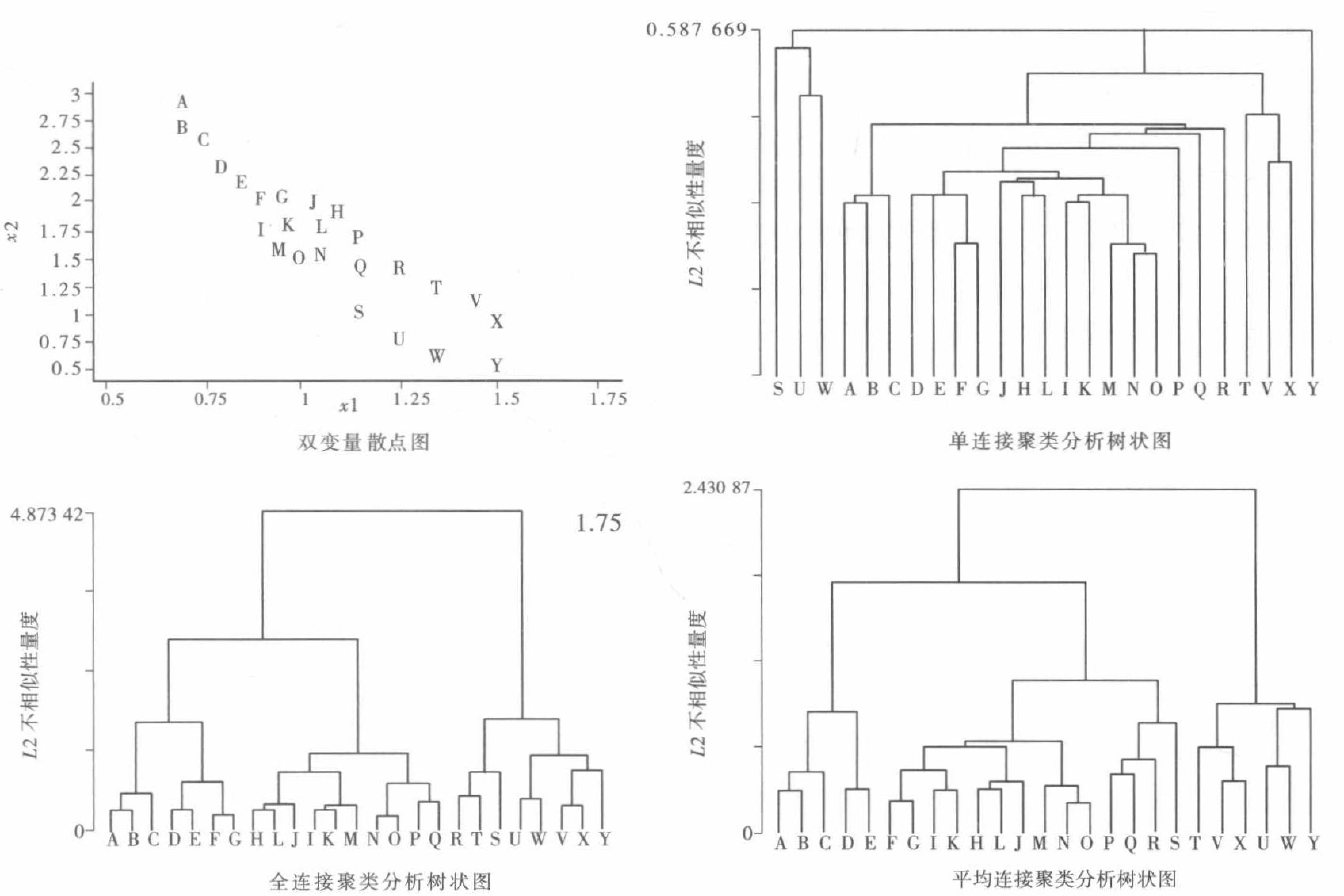



**图 1 普通数据集的三种聚类法**

图 1 的第一部分显示的是一张双变量散点图,共有可分类的单位 25 个。图中其余三部分都是使用欧几里得距离平方的聚类和树状图。对于这一例子而言,如果 $(x_i, y_i)$ 和 $(x_j, y_j)$ 是两个单位的坐标,那么它们之间的欧几里得距离的平方就是 $(x_i-x_j)^2+(y_i-y_j)^2$。右上方的部分显示的是一张单连接聚类的树状图。图 1 中右下和左下两张图分别是平均连接和全连接聚类的树状图。这些树状图的左边是不相似量度,水平线显示了被合并时的单位或类,而垂直线则一直在对类进行跟踪。平均和全连接聚类彼此最接近,二者都建议使用有三个聚类的聚类。然而,在类的细节方面二者则有所不同。单连接聚类与其他两种聚类明显不同。哪一种聚类"更好"可以通过考查聚类和散点图来加以判断——考查时要考虑单位是如何和为何会这样被分组(类)这一问题。而在实际分析时,则主要由研究的实际目的来引导这样的判断。

因为聚类这种工具是探索性的,所以不仅涉及聚类效度(步骤 4)的统计推论的工具为数不多,且它们的实用性也因为先验的探索而十分有限。从树状图选择类的数目常常被看作一个判断的问题,且任何一种聚类法都不难找到自己达到一个给定的聚类的独特理由。这一点也在图 1 中得到了阐述。

在聚类数已知或被假定时,有一些非分层的聚类方法可资利用。这包括 $k$-均值和 $k$-中位数聚类。另两种方法都采用首项和再定位(leader and relocation)算法。两种方法都是局部优化法,且为了避免到达局部最小,都必须重复许多次。定位算法是多雷恩等人(Doreian, Batagelj, & Ferligoj, 1994)开发的类似区块社会网络建模的直接聚类法的位置算法的核心。网络行动者和网络连接二者都被聚了类,而判别函数则据此根据实际关心的问题和网络【Network】连接被明确定义。正因如此,这样得到的聚类问题的解都不是特定的。

——Patrick Doreian

(沈崇麟译校)

## 参考文献


Doreian, P., Batagelj, V., & Ferligoj, A. (1994). Partitioning networks based on generalized concepts of equivalence. *Journal of Mathematical Sociology*, *19*, 1-27.

Everitt, B., Landau, S., & Leese, M. (2001). *Cluster analysis* (4th ed.). Oxford, UK: Oxford University Press.

Sokal, R., & Sneath, P. (1963). *Principles of taxonomy*. San Francisco: Freeman.


# 整群抽样(Cluster Sampling)

整群抽样是一种先将研究总体的单位分组,然后再选择一定数目的组的抽样方法。这些选出的组中的所有单位都在研究之列。这种抽样方法是多阶抽样【Multistage Sampling】法的一个特例。

——Peter Lynn

(沈崇麟译校)

# Cochran $Q$ 检验(Cochran's $Q$ Test)

威廉·科克伦(Cochran,1950)开发了配对或受试者内设计【Within-Subject Design】的 $Q$ 统计量。在这样的设计中,每个主体($r$)为每个实验条件($c$)提供一个二分的回答(dchotomous response)。Cochran $Q$ 检验跨所有条件的目标回答的概率是否相等。例如,如果 6 个主体($r=6$)进行 4 次"不及格"或"及格"检验($c=4$),Cochran $Q$ 可以评定跨四次检验的及格概率是否不同。当然,因变量可以是任何的二分变量,如及格-不及格、出席-缺席、击中-不中或增加-减少。

Cochran $Q$ 在数据以 $r$ 行 $c$ 列的矩阵形式排列时是很容易计算的。矩阵中的每一行代表一个含 $c$ 个元素的个案,元素的编码为"1"或"0",代表一个目标回答的存在或不存在。目标回答数跨行($u_i$)和列($T_j$)加总,并被用于在下面的公式中的 Cochran $Q$ 的计算:

$$Q=\frac{c(c-1)\sum_{j=1}^{c}(T_j-\bar{T})^2}{c(\sum_{i=1}^{r}u_i)-(\sum_{i=1}^{r}u_i^2)}$$

式中的 $T_j$ 是第 $j$ 列的合计,$u_i$ 是第 $i$ 行的合计,$\bar{T}=\frac{\sum_{j=1}^{c}T_j}{c}$。

Cochran $Q$ 服从 $c-1$ 个自由度【Degrees of Freedom】的 $\chi^2$ 分布。正如表 1 所示,来自每人参加了四次考试的 6 个主体的数据用 Cochran $Q$ 作了分析。然后将观察的 $Q$ 的统计值,和使用希望的第一类错误【Type Ⅰ Error】率(在这个例子中是 0.05)和自由度的 $\chi^2$ 值比较。在本例中 $Q$ 的统计值没有超过临界的 $\chi^2$。因此虚无假设【Null Hypothesis】被保留了,这意味着跨条件的考试及格的概率并无不同。

**表 1　Cochran $Q$ 的样本数据:复杂性变化水平检验**

| | 考试 1 | 考试 2 | 考试 3 | 考试 4 | |
|---|---|---|---|---|---|
| 主体 1 | 1 | 0 | 0 | 1 | $u_1=2$ |
| 主体 2 | 1 | 1 | 1 | 0 | $u_2=3$ |
| 主体 3 | 1 | 0 | 0 | 0 | $u_3=1$ |
| 主体 4 | 1 | 1 | 1 | 1 | $u_4=4$ |
| 主体 5 | 1 | 0 | 1 | 0 | $u_5=2$ |
| 主体 6 | 1 | 1 | 0 | 0 | $u_6=2$ |
| | $T_1=6$ | $T_2=3$ | $T_3=3$ | $T_4=2$ | $\bar{T}=\frac{\sum_{j=1}^{c}T_j}{c}=3.5$ |

$$Q=\frac{c(c-1)\sum_{j=1}^{c}(T_j-\bar{T})^2}{c(\sum_{i=1}^{y}u_i)-(\sum_{i=1}^{y}u_i^2)}=\frac{4(3)[(6-3.5)^2+(3-3.5)^2+(3-3.5)^2+(2-3.5)^2]}{4(2+3+1+4+2+2)-(2^2+3^2+1^2+4^2+2^2+2^2)}=6$$

$\chi^2_{(4-1)}=7.815$ 的第 95 百分位数

Cochran $Q$ 要求假定个案是从总体随机并独立地抽取的。它也要求协方差是齐性的,所有可能的列对的协方差假定是相等的。如果协方差齐性的假定被违反了,迈尔斯和韦尔(Myers & Well,1995)则推荐使用由格林豪斯和盖赛尔(Greenhouse & Geisser,1959)开发的 $\varepsilon$ 调整($\varepsilon$-adjustment)。修正后的 $Q$ 为:$Q^* = \varepsilon Q$。

调整检验的自由度为 $\varepsilon(c-1)$。与前面一样,$Q^*$ 依照调整自由度后计算的 $\chi^2$ 值来评估。$\varepsilon$ 的计算非常费事,因而在这里不再赘述。不过好在许多统计分析的程序在进行重复量度的方差分析【Analysis of Variance】时都会给出 $\varepsilon$ 的值。

在使用 Cochran $Q$ 时,样本量不宜小于16(上表中的样本量只是为了阐述方便)。跨条件显示,没有方差的(即全为 0 或 1)的那些个案对 $Q$ 没有贡献。重复量度的方差分析可看作 $Q$ 的一种备择方法,尽管它需要大样本来保证因变量抽样分布的正态性,且需要用 $\varepsilon$ 调整来修正协方差的异质性。

——James W. Griffith

(沈崇麟译校)

## 参考文献


Cochran, W. G. (1950). The comparison of percentages in matched samples. *Biometrika*, *37*, 256-266.

Greenhouse, G. W., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika*, *55*, 431-433.

Myers, J. L., & Well, A. D. (1995). *Research design and statistical analysis.* Hillsdale, NJ: Lawrence Erlbaum.


# 代码(Code)

参见编码【Coding】。

# 编码簿(Codebook)

参见编码框【Coding Frame】。

# 编码(Coding)

编码是用数字把语言资料转换成变量【Variable】和变量的类别的过程,以便把数据录入计算机进行分析。社会科学研究的数据【Data】是从自填式问卷【Self-Administered Questionnaire】、电话或面对面访谈的问卷、观察以及记录、文献、电影、图片或录音带中收集来的。从原本的"原始"形式来看,这些资料由语言表述、文字或视觉影像构成。尽管许多数据录入项目至少可以处理有限的语言材料,但就大多数分析而言,这些资料最终都必须转换成变量和数字,并输入计算机成为可读的数据集。一个

数据集中的每个变量（如性别）至少由两类组成，每一个都有一个独立的编码，这才成其为一个变量。它必须具有可变性；如果它不能改变，那么它就是常数。因此，一个研究可以包括男性和女性，这时性别（gender）或性（sex）就是一个变量，或者该研究也可以只包括男性或者女性，此时性别就是一个常数【Constant】。两种情况下都可以给性别设置数字编码。一个独特的编码可以由单位或多位数组成，有功能高数码（the functional upper number of digits）8 或 10 。

一个多位数的编码可在其中嵌套多个变量。例如在地震之后去询问被调查者是否无法获得各种公共设施，包括水、电、气、电话等。每种设施都可以生成一个二类变量，用 1 来代表“是的，断了”，2 来代表“不，没有断”。但这四个一位数的变量也可以组合成一个四位数的变量，把四个变量的各种可能性都包含其中。例如编码 1221，可以表明水和电话断了，但电和气还没断。这里指定的数字可以有其本身含有的意义，或者作为语言数据的速记表示。例如，在收集有关年龄的数据时，记录的每个人、机构、政府或国家、物件的数据都是有意义的，可以对这些人、国家或物件的相对年龄作一些陈述。相反，如果采集的是人、家庭或机构的宗教信仰的数据，指定给各种宗教的数字就是任意的——例如，1 代表天主教，2 代表佛教，3 代表印度教。

## 测量层次

在给语言资料设计编码时，经常会用到测量层次【Levels of Measurement】的概念（Stevens，1968）。有四种不同层次的测量：定类（nominal）、定序（ordinal）、定距（interval）和定比（ratio）。从定类数据到定比数据，在分析中可采用的计算和统计方法会变得更加多样和复杂。当为数据创建编码时，理解和考虑这些区别非常重要。

宗教编码被看作定类编码。定类编码是任意的、速记型类别标志符。在给定类变量生成编码时，研究者关心的是生成完备、互斥的编码。完备（exhaustive）表示给数据中变量可能出现的每一类取值设一个独立的编码数字。这样，在给宗教倾向变量生成编码或编码框的时候，可能就需要给每个没有宗教倾向的个人、机构和群体也生成一个独立的编码数字，那些被归为不可知论者、无神论者的人也需要独立的编码。互斥意味着每个人、机构、物体或群体的编码信息只能明白无疑地分配给变量的某一类。例如，佛教徒的编码只能是 2，不能是 1 或 3。

定类变量的类没有排序的意义，类和类之间的距离不能计算。对定类变量来说，唯一可以恰当计算的集中趋势【Central Tendency】量度是众数。许多社会科学最重要的变量都是自然的定类变量。包括性或性别、宗教、语言、种族认同和婚姻状态等。

定序变量的测量层级更高一级。定序变量的类别可有意义地顺序排列。因此，举例来说，一个喝了 1 品脱水的人所喝下的水少于喝了 2 夸脱水的人，而后者所喝的水又少于一个喝了 2 加仑水的人，这三个人可以分别被编码为 1，2，3。编码为 3 的人明显比编码为 1 和 2 的人喝的水要多，但在分析中，从这三个编码上无法判断此三人的相对距离。众数和中位数【Medians】都可以用于定序变量的计算，但在技术上讲，对得分进行加减或计算均值【Mean】都是错误的。给各种自然的定序变量进行编码的一个问题是设计既互斥又完备的编码。在上述例子中，编码是互斥的，但不是完备的。请注意这里没有对不喝水的人、喝 3 加仑的人或者 1 夸脱的人进行编码。

年龄可以被看作一个定距变量。定距变量需要有一个等距的（或恒常的）数字类

别尺度。试想一下这个问题的答案:“你去年过生日的时候几岁了?”对定距变量来说,它的类别既可以有序地排列,也可以说类别之间是等距离的。例如,定距变量都有假定零点,类别可以有序排列,而且我们还可以讨论类别之间的距离。比如说,24 和 26 之间的距离与 33 和 35 之间的距离相同。除了众数和中位数之外,诸如均值、标准差【Standard Deviation】等统计量和相关【Correlation】都是有意义的,可以被解释的。定比变量和定距变量的不同之处在于,它有一个有意义的零,而且是完全连续的,也就是说尺度上每个点都存在。质量和高度通常被看作定比变量。

定比变量也称作连续【Continuous】变量,而其他三类变量有时候则称作离散【Discrete】变量。对大多数社会科学研究而言,定距变量和定比变量之间的区别在很大程度上是任意的,因为定比变量可作的大多数统计分析也可以用于定距变量。相比之下,适合定距变量和定比变量的分析技术,在技术上不能被用于定类和定序变量。

## 确定变量的编码

数据的编码开发可以在研究设计阶段或者数据采集之后来做。当结构性问卷或抽象的形式用来搜集个数有限的变量数据,且大多数变量都有有限个可能的答案类别时,编码方案作为一种数据收集的工具,应该事先设计定稿。当观察【Observation】、焦点小组【Focus Group】、半结构化访谈【Semi-Structured Interview】或者文献【Document】都是数据来源时,一般在数据采集完成之后采用内容分析【Content Analysis】或其他程序,以归纳法来开发编码。我们把为问卷中的封闭式问题【Closed-Ended Question】的编码开发作为第一种例子,而将为开放式问题【Open-Ended Question】编码开发作为第二种例子。

开发编码时应该考虑许多问题。正如前面所强调的,编码和相关的答案类型应该既完备又互斥;也就是说,可能给出的每个答案应该都有一个编码,问卷的被调查者【Respondent】、数据的采集者以及数据的录入人员应该不难确定哪个答案符合所描述的情况,以及哪个数字或编码与这个答案相关。比如说,在洛马·普雷塔(Loma Prieta)地震之后,作者(Bourque & Russell, 1994; Bourque, Shoaf, & Nguyen, 1997)询问了被调查者,“你是否打开或者找到了一台电视机,或者收音机来获取有关地震的更多信息?”提供的答案包括“是的,普通的电视机。”“是的,用电池的电视机。”“是的,普通的收音机。”“是的,用电池的收音机。”以及“没有。”这五个答案预先分别编码为 1,2,3,4 和 5。调查者要求被调查者只选择一个选项作为答案。

尽管答案和编码的列表看似完备和互斥,实际并非如此。首先,许多被调查者从不止一种电子媒体获取信息;其次,许多被调查者也从汽车广播里获取信息。对这些问答的类别进行了修改,把汽车收音机加进了列表,而且也允许被调查者选择一个以上的答案。如果没有做这个工作,数据就有可能会丢失。

如果要记录或编码的多重信息在技术上属于一个变量,那么每个答案类别都成了一个有两个提及或未提及的编码或取值的变量。就这个例子来说,最初的一个变量与五个可能的答案就成了六个相关联的变量,每个都有两个可能的答案或编码。

其他类提供了另一个确保编码框【Code Frame】完备的办法。“其他”被用于处理研究者在设计编码框的时候没有预计到的答案。还是上述同一个研究,作者提供了有关房屋因地震受损的各种情况的列表,包括墙壁倒塌、屋顶垮塌等。作者没有预计到的是

加州圣克鲁斯城外山区居民所使用的水塔所遭到的损害。因为包括了“其他”这一类，当被调查者说有个水塔震坏了，访问者就把代表“其他”的编码19圈起来，在问卷中如下面的方式提供的空白处写下“水塔”如下：

其他 ………………………………………… 19

请具体说明：______________________

数据录入计算机的时候，表示预期之外的答案的编码19，以及对这个答案的文字描述“水塔”都被录入了数据集。

回顾以前的研究并对数据收集的过程进行前测【Pretesting】可以帮助研究者确定答案选项和相关的编码是否互斥与完备，但有时候社会变迁可能会导致经过充分测试的编码方案变得不准确、不完备，甚至完全过时。例如，婚姻状况曾经的标准分类是“从未结婚”“已婚”“离婚”“分居”和“丧偶”，但异性和同性伴侣同居的情况越来越多。根据2000年美国普查，美国5.2%的家户当时自称为“未婚伴侣家庭”。比如说，一个社会网络的研究，没考虑这个变化并对其作出解释，就可能会导致结果严重的失实。

确定用于收入、教育、年龄的类别和编码也是容易出问题的地方。一般来说，收入都被编码为定序变量。它会被分成如下的类别：低于10 000美元，10 000～19 999美元，20 000～29 999美元等。这样问题就来了——我们应如何确定每一类别的宽窄呢？最高和最低的类别又应该是多少呢？如果类别分得太宽，最高的类别设得太低，或最低的类别设得太高，就会出现堆积。将相当大比例的数据的码都编入单独一类，这种情况叫作堆积。例如，如果要研究的大学本科生，其年龄被当作定序变量分为18～21，22～24和25～29三类，其编码分别为1，2和3，绝大多数大学生的年龄都在18～21岁，他们就会被编码为1。这种情况出现的时候，变量就没什么意义了，因为它不再有任何方差【Variance】。

从那些基本结构本质上至少是定序的数据（如年龄、教育、收入等）来创建定序变量还会产生另一个问题，即可能会对数据分析有所限制。如果年龄被分成了18～24、25～29、30～34和35～39等几类，研究者虽然可以确认中位数年龄类，但却永远也无法计算平均年龄。不仅如此，在问卷上或资料收集表上写出年龄类别，比只是简单地询问上一次生日时的年龄和出生时间要占用更多的空间。这种情况下，资料收集表中设置一个简短的开放式问题，不仅能收集到数量和质量完全的数据，且生成的数据也更易于分析。

当研究者使用诸如“没有”“有些”和“很多”等分类的时候也会出现类似的问题。尽管没有的意思很清楚，但一些和很多是什么意思？它们又该如何编码？如果这个问题是要问被调查者认识的人中，有多少人在洪灾中受伤了，那么为什么不这样问：“你认识的人中一共有多少人在洪灾中受伤了？”这样收集来的信息可以生成近似于定距的数据，因而不会受制于人们对一些和很多这样一些词的各种解释。

在给数据编码的时候，记录丢失的信息和记录存在的信息一样重要。标准的缺失数据【Missing Data】的情况包括“不适用”“不知道”“拒绝回答”和“丢失”。当数据是直接问人问题采集来的时候，就会常常使用“不知道”和“拒绝回答”，而当数据来自记录或其他文件时，则“丢失”编码就极为重要。两种情况下都会用到“不适用”的编码，当有的数据与调查无关时，便可以采用“不适用”。之所以需要用到“不适用”，常常是因为需要编码的数据来自早前的信息，导致现在的数据与此无关。例如，不需要问男人他们怀孕了之后是什么感受；如果没有采血，也不需要在医疗记录中去找血红蛋白

的数据。

一般来说,如果要从文献或记录中提取数据,数据收集员会被告知要搜寻一些特殊种类的信息,并用编码加以记录。如果要给美国八大报纸一个月之内头版报道作一个内容分析【Content Analysis】,要求数据收集员记录下所有提到总统的文章和文章的"语气"(例如,是赞美的、批评的、分析性的等),而假如某报头版中没有提到总统,就是一个重要的信息。因此,表示头版报道内容的编码框或者数字编码的取值范围中就包含了为丢失所设的编码。代表丢失的编码至少提供了一个间接证明,说明数据收集员虽然寻找了这个有待编码的信息,但却没有找到。

在数据采集中,使用"不知道"这一编码的时机和程度是较有争议的问题。设置"不知道"这个类项及其编码的意义和价值因以下因素的不同而不同:研究者的目标、采集数据的方法、要搜寻事实性的信息还是有关态度的信息、是否告诉了被调查者可以选"不知道"这个类项,以及数据收集员受训练和督导的程度如何。如果数据是直接向人采集,主要的目标是收集事实性的信息或行为的信息,最好不要在数据收集中把"不知道"放在可选的类项中,因为有其他答案更合适的时候也会有调查员和被调查者忍不住去选"不知道"。更复杂的情况是,在诸如访谈等互动中,人们确实常常一边考虑问题,一边使用类似"不知道"这种短语。这种情况下,"不知道"表明这个人听到了问题,但不一定就是问题的答案。如果要问的是有关态度或者观点的问题,可能一个经过确认的"不知道",或者"没有意见"的答案类项是合适的,尤其是当被调查者被问及他们完全不了解的人、政策或做法的时候。

有些答案的类项在整个数据收集中重复出现,比如是、否、拒绝、不知道、信息丢失、不适用等。研究者可以在整个数据收集过程中给这些常见的答案选项使用一致性数字编码。表 1 展示了通常所采用的传统编码。

**表 1 一致性编码例示**

| 答案 | 编码范例 | | |
|---|---|---|---|
| | 一位数 | 两位数 | 三位数及以上 |
| 是 | 1 | | |
| 否 | 2 | | |
| 拒绝回答 | 7 | 97 | 997 |
| 不知道 | 8 | 98 | 998 |
| 丢失信息 | 9 | 99 | 999 |
| 不适用 | 0 | 00 | 000 |

最后,研究者应该考虑开发可以在数据分析中能将转换需求降到最低的编码框。比如说,如果要从健康儿童就诊(well-baby visits)的医学记录中收集有关免疫接种的数据,比起设计"没有接种""1~2 针""3~4 针"和"5 针以上"等类项,并用 1,2,3 和 4 来编码,记录每次就诊所做的免疫接种的真实次数要更有意义。

## 后编码

有些数据的编码框确实无法在数据收集程序定稿之时建好。未能将事先确定的答案类别清单包括在访谈中,其原因无非是研究者无法预估可能得到的数据的种类和数量,或者无法创建易于向数据收集员或数据录入人员说明并供其使用的完备且互斥的编码。在作探索性研究时情况更是如此,因为探索性研究的目标就是要查明再作更多的结构性研究是否可能及是否有价值。在这种研究中,访谈或者其他的数据收集方法可能是无结构的,或者只是部分有结构的。这些材料常常都是以音频或视频方式

记录的,并在编码和分析之前转录。当数据收集完成之后,任务就是要确定数据中存在什么模式(如果有任何模式的话)。这就是用诸如内容分析之类的程序进行归纳。分析文献、电影和电视节目也必须用到类似的程序。但最终的目标是构建出数据的编码框或数值表示,这样数据便可以被分析和概括。

内容分析及其相关的程序很复杂,关于如何操作的详细信息超出了本小节介绍的范围。不过这里可以介绍一些总的原则。首先,研究者必须设定编码框使用的具体目标,如果要观察正在玩耍的儿童,目标是测量行为的频次、行为持续的时间、互动的强度,还是某种组合?确认数据将要如何使用之后,研究者就可以确定哪些变量要抽取出来、哪些要被忽略掉,这个确认目标的过程可能涉及初步的总的内容分析,以便从数据中归纳发展出全部研究主题。

其次,在发展编码框的过程中,必须在细节过多和细节不足之间加以权衡。如果要描述衣服,那么给衣服上的纽扣数量和种类都进行编码是很重要的吗?或者就说这是件夹克就足够了?再次,与之相关的是,要在创建编码框的时候尽可能多地保存信息。这里特别关注的是那些只在一些文献中出现,但对要研究的问题特别重要的信息。因为内容分析是一个耗时费力的漫长过程,有时候因为研究者累了或者感到无聊了,就没有给重要的信息编码。只要原始数据还在,研究者总还是能重新检查数据,但这通常意味着看很多录音整理稿,或者听很多磁带。因此,在创建编码框的时候应尽一切努力把少见而重要的信息纳入其中。这就是积极搜寻数据,并对某种信息丢失的事实进行系统编码特别重要的原因所在。

最后,研究者需要给编码、变量或维度设置足够广泛的范围,好让编码员不必强行把数据塞入不合适的类项中去。比如说,以对下面这个陈述的回答为例:“描述这次交通事故中的交通工具。”首先,我们可以给这次交通事故中涉及的机动车数量进行编码。接下来,我们可以给事故涉及的机动车的种类进行编码——小轿车、运动型多功能车(SUV)或者小型货车。如果考查机动车的安全特性也是研究的目标之一,可能给车辆的品牌、使用年限、是否有安全气囊等进行编码也很重要。这里要说的重点就是一个刺激——无论是一个问题、文献中的一段话、一段时间的观察,或是一部电影或一盘磁带里的一个片段——都可能产生多种多样的信息,成为分析中的多个变量。

——Linda B. Bourque

(杨可译　沈崇麟校)

*也可参见定性数据编码【Coding Qualitative Data】。

## 参考文献


Bourque, L. B., & Clark, V. A. (1992). *Processing data: The survey example* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-085). Newbury Park, CA: Sage.

Bourque, L. B., & Russell, L. A. (with Krauss, G. L., Riopelle, D., Goltz, J. D., Greene, M., McAffee, S., & Nathe, S.) (1994, July). *Experiences during and responses to the Loma Prieta earthquake*. Oakland: Governor's Office of Emergency Services, State of California.

Bourque, L. B., Shoaf, K. I., & Nguyen, L. H. (1997). Survey research. *International Journal of Mass Emergencies and Disasters*, *15*, 71-101.

Stevens, S. S. (1968). Measurement, statistics, and the schemapiric view. *Science*, *161*, 849-861.

## 编码框(Coding Frame)

编码框、编码簿(codebook)或者编码手册是对如何将文字或视频资料转换成分析所需的数字数据的说明。它提供了文字资料和数字式数据之间的联系,并对数字数据的含义进行解释。

简单来说,编码框指的是在给代表单独一个变量【Variable】(例如人们出生的州)的信息编码【Coding】时使用的数字编码的取值范围。就这个例子而言,至少在编码框里要包括50个独特的数字,每个州对应一个数字。数字被任意地指派给各个州。对那些不是出生在美国的人和没有对应数据【Data】的人,则会设置特别的数字。一旦数据进入了数据文件,编码框就成了原始文字资料和数据集信息的位置和意义之间的唯一连接。没有编码框,分析员就不知道数字32代表什么,也可能不知道它表示一个州。

就最复杂的情况而言,编码框或编码手册说明了数据文件中所有数据的位置,每块数据或每个变量对应的是哪些数字集,每个独特数字代表着什么意思。因此,除了出生的州,数据集可能还包括了出生的县和出生日期。编码框可能会描述三个变量在数据集中出现的顺序,每个独特的编码代表什么,以及生日是以什么形式录入数据库的。

编码框可以在实际的数据收集之前建好,也可以之后通过内容分析【Content Analysis】和后编码过程创建。如果能提前知道变量的数量和每个变量可能收集到的数据范围,建议给变量作预编码,这样就可以在实际采集数据之前创建编码框。如果数据是通过计算机辅助的电话访谈(CATI)或类似方法采集的,大多数数据的编码框必须提前加规定或者预编码,以便为计算机编写采集信息的程序。CATI的一个主要优点在于访谈者可以在被调查者回答问题的时候就自动编码。如果访问的主要问题都需要访问员来打下冗长的文字答案,那么采用CATI的价值就小了很多。

有些种类的数据无法在数据收集之前就建好编码框,要么是因为可能获取的数据的种类和数量无法预测,要么是因为无法创建易于向数据收集员和数据录入人员说明并便于其使用的完备而互斥的编码。这种情况下,就必须在完成了数据收集之后通过归纳来创建编码框。这个完成的过程很复杂,无法在本小节一一介绍。

——Linda B. Bourque

(杨可译　沈崇麟校)

### 参考文献

Bourque, L. B., & Clark, V. A. (1992). *Processing data: The survey example* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-085). Newbury Park, CA: Sage.

## 定性数据编码(Coding Qualitative Data)

对很多不同类别的社会研究来说,编码【Coding】都是一个基本的组成部分。进行了数据收集之后,研究者就需要理解数据。编码要求研究者和数据进行互动,对其加以思考。编码从数据中创建出类项和概念,以一种系统的方式将数量巨大的数据集浓缩

成较小的可分析的单位。编码在位于不同部分但被认为具有相同特征的数据之间建立连接。编码便于组织、检索和解释数据，并进而在解释数据的基础上得出结论。

尽管定量研究和定性研究编码的过程可能很类似，因为两者都需要对数据进行分类拣选，但二者的编码目的却很不相同。例如，在给社会普查数据做编码的时候，可能会给某一类别指定其取值，以便统计处理（可能会进行频数统计）。相比之下，在半结构化访谈【Semi-Structured Interview】中所获取的类项并非规定的取值；而是从各个主题中挖掘出来的，仍然嵌入其情境。定性编码让数据得以确认、整理，并可以用不同的方式来加以考虑。

无论要编码的是问卷的开放式问题、访谈的录音整理稿还是田野笔记，编码的主要阶段就是要发展出一套可用于数据编码的类项。这些类项可能源自理论，也可能来自数据，可能从研究文献中来，或者以灵感为基础。斯特劳斯和科尔宾（Strauss & Corbin，1990）区别了三种不同层次的编码：开放的、轴向的以及选择性的编码。开放编码包括数据分解、比较和归类；轴向编码在开放编码后识别出来的类项之间建立连接，把数据放回一起；选择性的编码则要求选出核心的类项，将其与其他类项相关联，并证实和解释这种关系。如果我们设计研究项目的目的是检验假设，那么类项将要从所采用的理论框架中生成，数据就会被灌注进这些类项。这种编码叫作向下编码（coding down）。如果数据要用于生成理论，类项就从数据中生成，这就涉及向上编码（coding up）（Fielding，1993）。

科菲和阿特金森（Coffey & Atkinson，1996）强调指出，编码既可以用于数据简化，也可以用于数据复杂化，不过通常编码都是这二者的组合。编码可以把数据简化成简单而宽泛的类项或共同点，以便于检索具有共同特征的数据。诚如斯特劳斯（Strauss，1987）所示，编码同样可以用于对以前没有考虑过的分析的可能性加以发展、重新阐释并开始分析，帮助生成理论。

过去的十年里出现了数不清的电脑软件包，可专门用于简化定性数据编码的过程。尽管计算机定性数据编码快速而又简便，可以使用大量的编码，并开启意想不到的探索和分析之路，但它给研究者带来的处理文字资料的方式实在是太多了。不仅如此，研究者也可能会受制于计算机程序。它可能会对分析带来难以接受的限制，而且定性编码也可能会受形式化的支配。

——Sharon Lockyer

（杨可译　沈崇麟校）

## 参考文献


Coffey, A., & Atkinson, P. (1996). *Making sense of qualitative data: Complementary research strategies.* Thousand Oaks, CA: Sage.

Fielding, J. (1993). Coding and managing data. In N. Gilbert (Ed.), *Researching social life* (pp. 218-238). Thousand Oaks, CA: Sage.

Strauss, A. (1987). *Qualitative analysis for social scientists.* Cambridge, UK: Cambridge University Press.

Strauss, A., & Corbin, J. (1990). *Basics of qualitative research: Grounded theory procedures and techniques.* Newbury Park, CA: Sage.


# 系数（Coefficient）

系数是关系【Relationship】的数值。一个系数一旦被估定，便表明了两个或更多个

变量之间的联系。例如,变量 $X$ 和 $Y$ 之间的皮尔逊相关系数【Pearson Correlation Coefficient】0.90 这一数值,精确地把二者的关系数量化,表示二者的关系很强。在有着两个自变量 $X$ 和 $Z$ 的多元回归【Multiple Regression】方程中,$X$ 的 6.3 的回归系数表明,在保持 $Z$ 不变时,$X$ 的一个单位的变化,可期望 $Y$ 有 6.3 个单位的变化。一般来讲,对于相关的变量而言,可资利用的系数有许多个,每一个都有不同的性质,而在有些时候,究竟什么性质才是我们所愿,则要取决于分析者的目的。

——Michael S. Lewis-Beck

(沈崇麟译校)

*也可参见参数【Parameter】。

## 决定系数(Coefficient of Determination)

决定系数是普通最小二乘线性回归【Ordinary Least Squares Linear Regression】的拟合优度量度【Goodness of Fit】(在多变量中它常被叫作复决定系数)。双变量的决定系数用 $r^2$ 表示,而多变量的决定系数则用 $R^2$ 表示。这一量度表示估计的回归线在多大程度上与实际的数据拟合。它告诉我们因变量 $y$ 的变化在多大程度上是可以为把所有的自变量【Independent Variable】放在一起的模型所解释。特别要指出是,它是 $y$ 的总变化中,$y$ 被解释的变化的比例。

有若干种公式可用来计算决定系数。第一种是先取被解释的平方和,然后再将它除以总平方和:$R^2=\text{ESS}/\text{TSS}$。第二种是用 1 减去残余平方和在总平方和中的比例(未解释的平方和):$R^2=1-(\text{RSS}/\text{TSS})$。

因为它是解释的平方和在总平方和中的比例,所以这一量度的值一定是 0~1。1 这一值表示回归线与数据完全拟合,在值接近 0 时,表示拟合得很差。如果 $x$ 和 $y$ 完全线性独立,那么 $R^2$ 便将等于 0。必须注意的是,如果样本平均数解释的因变量的变化多于自变量所解释的,那么 $R^2$ 的值有可能是负的,值域为 0~−1。将量度值乘以 100 得到的值,使我们得以更清楚地解释自变量的被解释的变化。用这样的计算,对一个 0.25的 $R^2$,我们可以作这样的解释:因变量变化的 25%可为模型中的自变量所解释。

决定系数的平方根叫作复相关【Multiple Correlation】系数,而在双变量中则称为样本相关系数。相反,$R^2$ 则是复相关系数的平方,它是从自变量计算得到的估计的因变量($\hat{y}$)和实际的因变量($y$)之间的相关量度。

尽管这一量度要考虑的问题有很多,有时也许还会考虑过多,但其中有些要点则是在对这一量度作解释时必须记住的。第一,不论你的 $R^2$ 有多么高,它也只不过是相关的证据,都不能对因果关系提供正面的支持。这就是说,即使你的 $R^2$ 很高,你也不能说是你的自变量引起了你的因变量的变化这样的话。第二,虽然 30%不是一个非常高的值,但你的模型可能已经运行得比较好了;解释了政治环境中的一个因子的变化的 30%,这一比例常常已经有了很重要的意义。此外,在进行时间序列(Time-Series)分析时,我们常常会发现有些 $R^2$ 会在 0.90 以上。第三,当将自变量加进模型时,$R^2$ 会随之增加,但对那些理论取向的模型,我们应该避免试图将相关系数最大化的做法。最后,不能对任何两个给定的模型的 $R^2$ 进行比较,如果它们没有相同的因变量

的话。

在使用多变量模型时,一种行之有效的做法是考虑使用调整【Adjusted】的 $R^2$,因为每一个加入模型的自变量通常都会增加被解释的变化的数量,所以考虑一下自由度【Degrees of Freedom】问题是很重要的。而调整的 $R^2$ 正是出于这一考虑而对那些增加的参数进行调整的。

——Christine Mahoney

(沈崇麟译校)

## 参考文献


Frankfort-Nachmias, C., & Nachmias, D. (2000). *Research methods in the social sciences* (6th ed.). New York: Worth.

Gujarati, D. (1995). *Basic econometrics* (3rd ed.). New York: McGraw-Hill.

Wooldridge, J. M. (2000). *Introductory economics: A modern approach*. Cincinnati, OH: South-Western College Publishing.


# 认知人类学(Cognitive Anthropology)

认知人类学是以社会关系为背景对人类思想所进行的研究。有些认知人类学的著作考察了离散的认知过程,例如记忆或者直觉,但在古迪纳夫(Goodenough,1957)将文化定义为个人作为社会群体的成员行动得当所必须掌握的知识之后,现代认知人类学的发展才打下了良好的基础。在这个看似简单的定义中暗含了一些有关知识的分享和分布的问题,这些知识现在已成为该学科的核心领域。认知人类学一直是人类学中最具方法论自觉的分支学科,这一定程度上也是拜这些有关分享和分布的问题所赐。许多这种工作都围绕着寻找某个社会群体(无论是一个社会网络、社会组织,还是一个社区)的文化上凸显的范畴,以及对组织这些范畴的基本要素的测定来进行。

调查者用于探求那些标志类别的词语,以及调查者用来定义这些词语的含义的区别性特征的技术有如下几种:自由列表(free-listing),借此可以生成组成某个文化领域的词语;各种形式的非强制性或强制性堆积分类(pile sorts),由此可以确定各个词语的意义的异同点;各种形式的评级和排序工作,以精确测量意义的不同维度的重要性(参见 Weller & Romney, 1988)。罗姆尼、韦勒和巴彻尔德(Romney, Weller, & Batchelder,1986)提出的文化共识模型给考察文化知识的分享和分布带来了巨大的发展。文化共识模型用被调查者之间的相关关系来测量共识;如果这种被调查者之间的关联可以用单独一个主因子来解释,那么就存在共识,而且也可以因此而合理地推论,所有的被试者都信赖那个特别文化领域的一个单独的、共享文化模型。围绕这一共识的被调查者的分布(包括备择的替代性文化模型)以及这一共识的具体内容也可以得到检视。

不同民族背景下的内科医生和病人之间的生殖系统癌症文化模型的研究是这种路数的一个范例(Chavez, Hubbell, McMullin, Marinez, & Mishra, 1995)。上面所描述的工具被用来考察南加州的医生、英裔妇女以及本地拉丁裔和拉丁移民之间的癌症文化模型。在每个群体内部都有很明显的共识(也就是说,群体内部彼此之间对导致生殖系统癌症的重要原因的看法一致)。在英裔妇女与医生、英裔妇女与本地拉丁裔、本地拉丁裔与拉丁移民之间都存在一些交叠的共识。但是拉丁移民却与医生和英裔妇女之间共享的癌症原因模型相距

甚远。这些结果可以用图来表示,展示出个人和群体在文化意义空间上是如何分布的。

认知人类学的这种系统的民族学技术加强了传统的文化人类学定性技术。人类学研究的一个主要目标就是理解其他人是如何看待世界的。在资料收集和分析上的方法论革新有助于研究者更准确地理解他者的文化模型。

——William W. Dressler

(杨可译　沈崇麟校)

## 参考文献


Chavez, L., Hubbell, F. A., McMullin, J. M., Marinez, R. G., & Mishra, S. I. (1995). Structure and meaning in models of breast and cervical cancer risk factors: A comparison of perceptions among Latinas, Anglo women and physicians. *Medical Anthropology Quarterly*, *9*, 40-74.

Goodenough, W. (1957). Cultural anthropology and linguistics. In P. L. Garvin (Ed.), *Report of the 7th Annual Roundtable on Linguistics and Language Study* (pp. 167-173). Washington, DC: Georgetown University Press.

Romney, A. K., Weller, S. C., & Batchelder, W. (1986). Culture as consensus: A theory of culture and informant accuracy. *American Anthropologist*, *88*, 313-338.

Weller, S. C., & Romney, A. K. (1988). *Systematic data collection* (Qualitative Research Methods, Vol. 10). Newbury Park, CA: Sage.


# 同期群分析(Cohort Analysis)

同期群分析是一种考察数据的一般性策略,而非一种统计方法。在过去的数十年间,特别是在20世纪最后的25年间,它在社会科学中的使用变得越来越普遍。它在评估(人或其他存在的)老龄化的后果及理解社会和文化的变迁中很有用。同期群分析之所以引起人们的注意,也在于它在方法学上具有异乎寻常的激起人们兴趣的挑战性。

就其最广泛的意义而言,凡使用同期群概念的量度,并将这一量度与一个或更多个附加的变量联系起来的量化研究,都可称为同期群分析。一个同期群依次包含的那些个体(人或其他个体),在一个特定的时间经历了一个特定的事件。常常被研究的那种同期群是人的出生年同期群,即那些在给定的年、十年或其他时期出生的人。如果同期群这一词不带修饰词那么它通常都指一个人的出生同期群。否则会用个体共同经历的事件作为修饰词来确定同期群的种类,如退休同期群。用于定义同期群的事件涵盖从结婚到参加一个组织,从参加研究生课程学习到第一次为人父(母)这样的各种事件。个体如果不是人类则可以是结婚、组织、教科书、电影或其他在特定期间形成的存在。

通常和更为严格的同期群分析的含义是指那些任何试图估计年龄、时期和同期群对因变量如某种行为或态度的效应,且分析至少涉及来自两个同期群的至少两个时点的数据那样的分析。年龄效应起因于成长、同期群效应起因于同期群的成员资格,而时期效应则起因于随时间变化而来的影响。例如,处在青少年或青年时期的人都酷爱通俗音乐就是一种同期群效应,而由经济萧条引起的态度变化就是一种时期效应。显而易见,在老龄研究中估计年龄的效应是很重要的,而在理解社会和文化变迁时,对诸如

这样的效应进行估计同样也是很重要的。

## 年龄、时期和同期群难题

年龄、时期和同期群效应的估计并非是直截了当的,因为在来自标准的统计分析的发现中,两种效应是彼此混杂在一起的。例如,在某一时点的不同年龄的人之间的差异,可能是由于年龄效应、同期群效应所致,或两种效应共同所致。无独有偶,在追踪调查不同时期的人的研究中观察到的变化,可能是年龄效应、时期效应所致,或二者共同所致。效应混杂是识别问题【Identification Problem】的一个例子,它多见于在三个或更多个自变量可能对一个因变量产生影响,而每个自变量又是其他自变量一个精确的线性函数时。年龄是时期和同期群的一个线性函数,同期群是年龄和时期的线性函数,而时期又是年龄和同期群的线性函数。

有一种启发式的图表常被用来帮助我们厘清同期群数据中的效应混杂,它就是标准同期群表,该表中有并列的两组或更多组的某个与年龄相关的截面数据,各组的年龄都以相等的间隔分类,如表1的年龄数据都以十年为间隔分类。该表行中的时期和同期群的效应是混杂的,而年列中的年龄和同期群效应也是混杂的。期群效应可通过时间来显示,即通过自上向下和自左向右的对角线上的数字显示,而年龄和时期的效应在同期群对角线上是混杂的。

**表1 1974、1984和1993年全美综合调查(American General Social Surveys),被调查人认为婚外性行为总是不对的人数的百分比**

| 年 龄 | 1974年 | 1984年 | 1994年 |
|---|---|---|---|
| | n/% | n/% | n/% |
| 20~29 | 59.2(392) | 68.3(384) | 83.8(346) |
| 30~39 | 70.9(291) | 63.9(304) | 77.7(462) |
| 40~49 | 75.3(270) | 70.8(241) | 76.7(411) |
| 50~59 | 80.8(278) | 78.6(181) | 77.2(272) |
| 60~69 | 83.0(194) | 79.7(166) | 85.8(165) |

注:为了增加代表性,数据以户中18岁及以上人数作了加权。

年龄、时期和同期群的线性效应,以没有一种统计分析可以把它们区分开来的方式混杂在一起,无限个年龄、时期和同期群效应的组合才可以对跨年龄、时期或同期群对一个因变量的线性变化模式作出解释。有一种方法可以将非线性效应分开,但这只有在效应是可加的和可以对某些效应作出精准的假设时才能做到,而这些条件几乎是无法满足的。

然而,即使没有做严格的模型检验,也常常仍有可能对效应的一般性质作一个合理的结论。某些有用的同期群分析法将理论、手头数据之外有关正在研究的现象的信息、常识、简单的统计分析和正式的数据考察结合在一起,得出虽然是初步的,但却是站得住的结论。

## 一个例子

表1的数据便可用来作为对描述性数据进行简单的效用考察的例子。1974年的数据显示,年龄与对婚外性关系的限制性态度之间存在正相关关系,而同期群对角线则表明,总的来讲,每一个出生于同期群的人,随着年龄的增长,其限制性态度也有所增长。这些数据本身说明在限制性态度上具有正的年龄效应。然而在更为仔细地考察表中的数据之后,我们便会对这样的解释产生怀疑。在20~29岁这一个年龄组,我们可以清楚地看到时期对限制性的影响,可能已经强到足够使同期群内的限制性态度有所

增长。根据理论,同时也得到许多经验证据的支持,年轻人较之老年人更容易受到时期的影响,因此受早期性解放思想影响的年龄差异可以解释 1974 年数据存在的年龄和对婚前行为约束的态度的差异。由此可见,数据中并不存在任何令人信服的年龄效应的证据。

## 年龄、时期和同期群特征模型

罗伯特·布里安和他的合作者引进了一种新的分析方式,尽管它提供不了有意义的年龄、时期和总同期群效应的估计值,但我们有理由把它视为一种同期群分析。虽然它无法让相互联系的变量中的两个保持不变的同时,让第三个变化,但是它可以设法使年龄和时期不变,而令同期群的某种特征,例如它们的相对大小变化。如果同期群特征不承载对同期群很强的线性关系,那么它的效应估计值应当是有意义的,即使模型无法估计年龄和时期效应的值。但已经有证据表明,这种很有前途的方法在研究诸如犯罪和受害这样的因变量中是很有用的。

——Norval D. Glenn

(沈崇麟译校)

## 参考文献

Glenn, N. D. (2003). Distinguishing age, period, and cohort effects. In J. Mortimer & M. Shanahan (Eds). *Handbook of the life course* (pp.465-476). New York: Kluwer Academic/Plenum.

Mason, K. O., Mason, W. M., Winsborough, H. H., & Poole, W. K. (1973). Some methodological issues in the cohort analysis of archival data. *American Sociological Review*, *38*, 242-258.

O'Brien, R. M. (2000). Age-period-cohort-characteristic models. *Social Science Research*, *29*, 123-139.

# 共整合(Cointegration)

当一个不平稳的变量线性组合是平稳的时候,那么我们就说这些变量被共整合了,定义那个平稳的线性组合的向量叫作共整合向量。一个时间序列是平稳的,如果它的分布不随时间变化而变化的话。最简单的平稳过程的例子是$\{\varepsilon_t\} = \cdots, \varepsilon_{-1}, \varepsilon_0, \varepsilon_1, \cdots$,它们表示一系列独立的和恒分布的随机变量。下标 $t$ 指的是时间。如果变量的分布取决于时间 $t$,那么它就是不平稳的。在共整合分析中,最常见的非平稳性形式是整合变量(integrated variable)的非平稳性。随机游走【Random Walk】,$X_t = X_{t-1} + \varepsilon_t = X_0 + \sum_{i=1}^{t} \varepsilon_i$是一个一阶整合非稳定变量的例子。

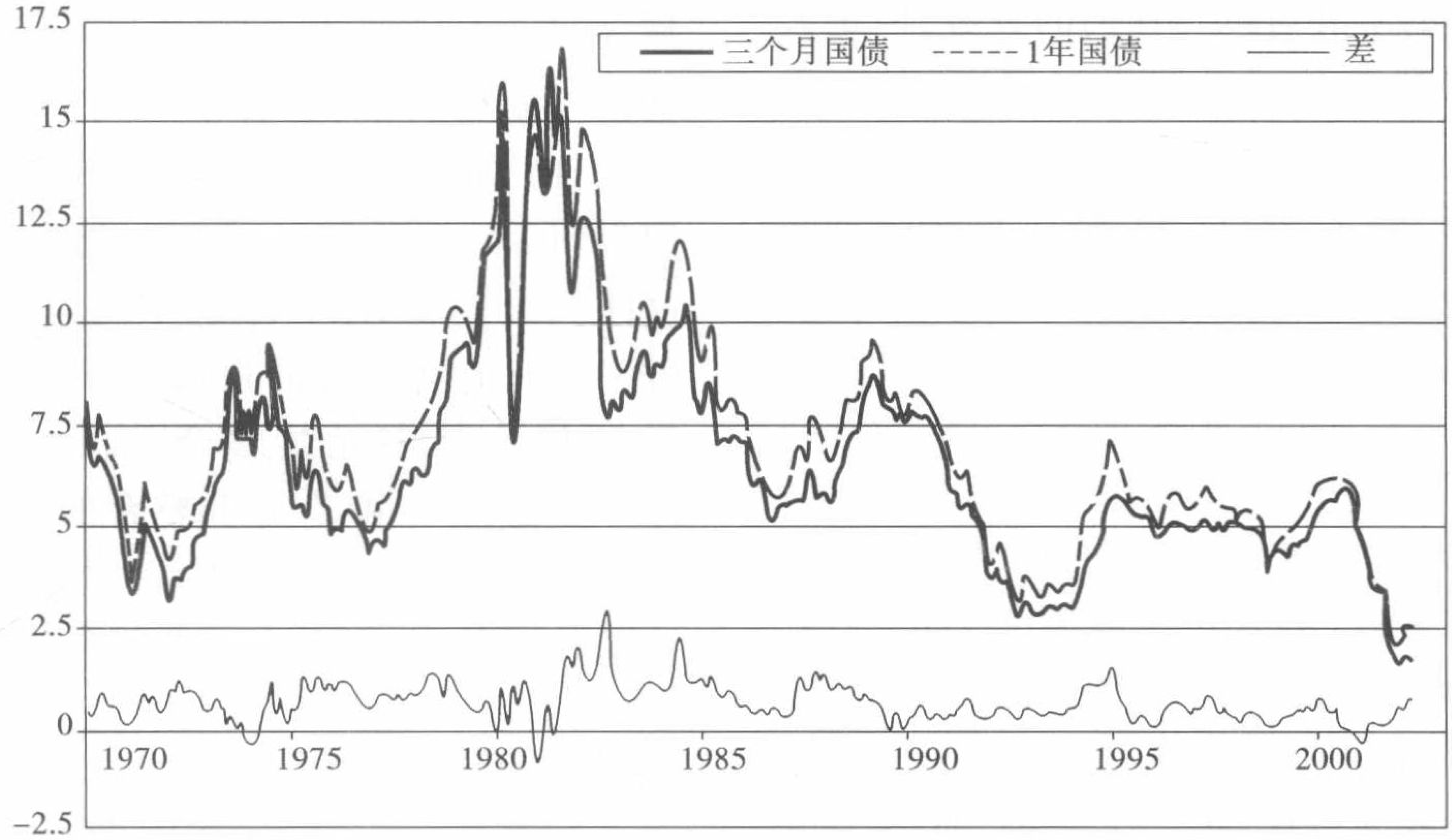
共整合的概念之所以令人感兴趣,是因为它能用于揭示可能有理论解释的关系。例如,共整合关系可能由一个经济模型的一阶条件定义,而共整合分析则可被用来估计经济模型和检验某些理论假设。

共整合的另一个经典例子是有关一条跟着它喝醉了的主人的狗的。在这个例子中,狗和它的主人的位置,作为时间的函数,是两个非平稳过程,因为喝醉的主人四处走动,步履的方向是随机的。然而,这两个过程是共整合的,因为狗和主人之间的距离是平稳的。

另一个例子是有着不同期限的债券的利率的性质。单独的利率可能都有很大的波动,但是它们之间的差似乎是平稳的。这

一点在表 1 中得到了阐释。该图列有从 1970 年 1 月—2002 年 4 月期间的三个月的国债和一年的国债的利率。哪根细实线是两种利率的差？

表 1 三个月的短期国债和一年的长期国债的利率(从 1970 年 1 月—2002 年 4 月)




注：两种利率似乎是非平稳的，而它们之间的差好像是平稳的，所以这两种利率是共整合的。(数据来源：圣路易斯联邦储备银行)

## 历史发展

共整合是由克里弗·格兰杰(Granger, 1981, 1983)引进的，而共整合过程的统计分析则由罗伯特·恩格尔和格兰杰(Engle & Granger, 1987)、苏林·约翰森(Johansen, 1988, 1991)及菲利普斯·彼得(Phillips, 1991)正规化的。共整合涉及许多计量经济学【Econometrics】的概念，如单位根过程、伪回归、共同随机趋势(有关这方面的出色评论，请参见 Watson, 1994)。现在有许多有关共整合和共整合的应用的文献。近年来的发展则主要在改进和概括现有的技术，如偏倚及 Bartlett 修正、微共整合过程、周期共整合、追踪组共整合(panel cointegration)、非线性共整合和关系结构性变迁过程的共整合。许多参考文献都可以在詹姆斯·汉密尔顿(Hamilton, 1994)的著作中找到。他的著作也包含了非常好的共整合技术的导论。有关共整合的似然分析的优秀教科书是由约翰森(Johansen, 1995)编撰的，它的姐妹篇则是由彼得·汉森和约翰森(Hansen & Johansen, 1998)所撰。

## 整合的阶

那些要共整合的变量都必须进行个别的整合。正规地说，譬如说，一个时间序列 $X_t$，要被整合成 $d$ 阶，如果 $(1-L)^dX_t=\psi(L)\varepsilon_t$ 是非平稳的，且 $\psi(1)=\sum i\psi_i\neq0$ 的话，式中的 $L$ 是滞后算子，如 $\psi(L)\varepsilon_t=\psi_0\varepsilon_t+\psi_1\varepsilon_{t-1}+\cdots$ 那样，记号 $X_t\sim I(d)$ 是“$X_t$ 是 $d$ 阶整合”的简写，而记号 $\Delta^dX_t$ 有时则被用来替代 $(1-L)^dX_t$。整合的阶的一种备择的定义假定 $\{\varepsilon_t\}$ 是一种白噪声【White Noise】过程，并用协方差平稳代替平稳。白噪声过程定义如下：如果有期望值 0，$E(\varepsilon_t)=0$ 和不随时间

变化的有限方差 $\mathrm{var}(\varepsilon_t)=\sigma^2$，且跨时间不相关 $s\neq t$，$\mathrm{cov}(\varepsilon_s, \varepsilon_t)=0$。一个协方差平稳过程，$\{X_t\}$有常均值和方差及一个只依赖于时间差的自协方差函数，$\mathrm{cov}(X_s, X_t)=\varphi(|s-t|)$。

整合的阶 $d$ 可以是任何的真数，但是大多数文献都涉及了单位根过程，它对应于 $d\in\mathbb{N}$（正整数）。非整数被称为微整合过程，它与长记忆性过程（long-memory processes）有关。

一个变量的向量整合的阶，$X_t=(X_{1t}, \cdots, X_{pt})'$，被定义为单个元素 $X_{1t}, \cdots, X_{pt}$ 整合的最高阶。

如果 $X_t\sim I(d_x)$ 和 $Y_t\sim I(d_y)$，且 $a$ 和 $b$ 是非零的常数，那么在一般情况下 $aX_t+bY_t$ 是 $d=\max(d_x, d_y)$ 阶的被整合。在 $aX_t+bY_t$ 是以小于 $\max(d_x, d_y)$ 的阶被整合时，它描述了共整合的特征，是一个特例，对两个被共整合的变量而言，它们必须以相同的阶整合。记号 $CI(d,b)$ 被用于那些有共整合关系的变量的 $d$ 阶整合。它是 $d-b$ 阶的整合。

共整合更为广义的概念是多元共整合（multi-cointegration），有时它也叫作多项共整合（polynomial cointegration）。如果 $(1-L)^d X_t$ 和 $Y_t$ 以某个 $d\neq 0$ 的阶共整合，那么 $X_t$ 和 $Y_t$ 就是多元共整合的。例如，$X_t\sim I(2)$ 和 $Y_t\sim I(1)$，那么 $\Delta X_t=X_t-X_{t-1}$ 就可能和 $Y_t$ 共整合。

## 实例

令$\{\varepsilon_t\}$为一独立的和恒等分布的变量系列，且假设对于 $t=1,2,\cdots$

$$X_t=\sum_{s=1}^{t}\sum_{i=1}^{s}\varepsilon_i+\sum_{s=1}^{t}\varepsilon_s+\varepsilon$$

$$Y_t=2\sum_{s=1}^{t}\sum_{i=1}^{s}\varepsilon_i+\varepsilon_t$$

$$Z_t=-\sum_{s=1}^{t}\varepsilon_s+\varepsilon_t$$

因为 $(1-L)^2X_t=3\varepsilon_t-3\varepsilon_{t-1}+\varepsilon_{t-2}$ 是平稳的，且系数满足 $\sum_i\psi_i=3-3+1=1\neq 0$，我们看到 $X_t$ 是二阶被整合的。同样，它服从 $Y_t\sim I(2)$ 和 $Z_t\sim I(1)$，这样向量 $(X_t, T_t, Z_t)'$ 也是二阶被整合的。不仅如此，$2X_t-Y_t=2\sum_{s=1}^{t}\varepsilon_s+\varepsilon_t$ 则是一阶整合的，这样 $X_t$ 和 $Y_t$ 被共整合，$CI(2,1)$，有共整合向量 $(2,-1)'$。同样，我们看到 $2X_t-Y_t+2Z_t=3\varepsilon_t$ 是零阶被整合的，这样三个变量，$X_t$，$Y_t$ 和 $Z_t$ 是共整合的，$CI(2,2)$，有共整合向量 $(2,-1,2)'$。$X_t$ 和 $Z_t$ 同样也是多元共整合的，因为这两个 $I(1)$ 变量，$\Delta X_t$ 和 $Z_t$ 是共整合的，$CI(1,1)$，因为 $\Delta X_t+Z_t=3\varepsilon_t-\varepsilon_{t-1}$ 是 $I(0)$。同样，我们可以看到 $Y_t$ 和 $Z_t$ 是多元共整合的。

## 共整合过程分析

整合变量的存在违反线性回归模型中的经典假定，主要是因为非平稳性。这可能会引起无法得到渐近正态的参数估计量，致使共整合变量的统计分析比标准的回归分析更为复杂。

用于共整合的两种主要路数是基于回归的路数和似然路数。前者涉及带有各种修正的标准回归技术，而后者则基于向量自回归模型（VAR），有时它也叫作 Johansen 法。变量系统的共整合分析一般由以下步骤组成：

- 确定共整合的秩。这一步一般都要涉及 Dickey-Fuller 型分布的使用。这些分布有很复杂的表达式，涉及布朗运动的随机积分。幸运的是，最现代的计量经济学的教科书，包括汉密尔顿（Hamilton，1994）编写的那一本，都包含有最为普遍的 Dickey-Fuller 分布表。约翰森（Johansen，1988，1991）的轨迹检验是一种比较流行的用于确定共整合秩的检验。
- 估计共整合关系和其他参数。没有其他的限制条件/正态化，就无法确认共整

合关系。

• 对共整合参数和其他参数进行推论。对共整合关系进行分析特别有用,因为它们也许可以被理解为经济模型中的结构方程。后者意味着小样本也可以有有关共整合关系非常丰富的信息。在标准情况下,其他参数的推论要遵循传统的结果。

约翰森(Johansen,1988)的似然分析是以VAR模型为基础的,$X_t = \Pi_1 X_{t-1} + \cdots + \Pi_k X_{t-k} + \varepsilon_t$,式中的$X_t$是一个$p$维度向量,而$\{\varepsilon_t\}$则是一个均值为0和协方差矩阵为$\Omega$的独立正态分布变量系列。这个模型以误差修正模型(ECM)的形式被重写为:

$$\Delta X_t = \Pi X_{t-1} + \Gamma_1 \Delta X_{t-1} + \cdots + \Gamma_{k-1} \Delta X_{t-k+1} + \varepsilon_t$$
$$t = 1, \cdots, T$$

如果$X_t$的元素是共整合的,那么$p \times p$矩阵$\notin \Pi$,有约化的秩$r < p$,而$X_t$中的公共随机趋势数则由$p-r$给出。$\Pi$的约化秩可由$\Pi = \alpha\beta$将该矩阵表达为两个矩阵的积求得,式中的$\alpha$和$\beta$都是$p \times r$矩阵。这样的表达法将导出可以用约化的秩回归估计的回归方程。一种分析共整合过程的重要工具是Granger表述法,$X_t = C\sum_{i=1}^{t}\varepsilon_i + C(L)\varepsilon_t + X_0 - C(L)\varepsilon_0$。这种表述法把这一过程分解成非平稳的随机游动,$C\sum_{i=1}^{t}\varepsilon_i$;平稳成分$C(L)\varepsilon_t$;以及一个依赖于初始值的项$X_0 - C(L)\varepsilon_0$。一个重要的结果是$\beta' C = 0$(一个$r \times p$的零矩阵),由此而导出$\beta' X_t = \beta' C(L) \sim I(0)$。这样共整合关系便由$\beta$的列给出。关系$\Delta X_t = \alpha(\beta' \overline{X}_{t-1}) + \cdots$显示了$\alpha$定义$X_t$如何在$\beta' X_{t-1}$中对它的期望值的离差作出反应,因此$\alpha$有时也被叫作调整的系数矩阵。$\beta' X_{t-1}$有时也被解释为对均衡的离差。在这里,$\alpha$给我们提供了朝均衡关系调整的变量的信息,即变量朝均衡调整的方向和收敛的速度的信息。

——Peter Reinhard Hansen

(沈崇麟译校)

## 参考文献


Engle, R. F., & Granger, C. W. J. (1987). Co-integration and error correction: Representation, estimation and testing. *Econometrica*, *55*, 251-276.

Granger, C. W. J. (1981).Some properties of time series data and their use in econometric models specification. *Journal of Econometrics*, *16*, 121-130.

Granger, C. W. J. (1983). *Co-integrated variables and error-correcting models*. Discussion Paper No. 1983-13, University of California, San Diego.

Hamilton, J. D. (1994). *Time series analysis*. Princeton, NJ: Princeton University Press.

Hansen, P. R., & Johansen, S. (1998). *Workbook on cointegration*. Oxford, UK: Oxford University Press.

Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control*, *12*, 231-254.

Johansen, S. (1991). Estimation and hypotheses testing of cointegrating vectors in Gaussian vector autoregressive models. *Econometrica*, *59*, 1551-1580.

Johansen, S. (1995). *Likelihood-based inference in cointegrated vector autoregressive models*. Oxford, UK: Oxford University Press.

Phillips, P. C. B. (1991). Optimal inference in cointegrated systems. *Econometrica*, *59*, 283-306.

Watson, M. W. (1994). Vector autoregression and cointegration. In R. F. Engle & D. L. McFadden (Eds), *Handbook of econometrics* (Vol. 4, pp. 2843-2915).Amsterdam: Elsevier.


# 共线性(Collinearity)

参见多重共线性【Multicollinearity】。

## 列关联(Column Association)

参见关联模型【Association Model】。

## 共同性分析(Commonality Analysis)

共同性分析是一种用于将解释/预测的方差,无论它是在一个测量的变量,还是在一个潜变量中,分解成为(a)每一测量的预测变量【Predictor Variable】单独解释的和(b)预测变量的每一可能组合共同解释的子成分。解释的方差和所有解释的方差的子成分,每一个都是被解释的方差的度量的平方。共同性分析原来是被开发出来用于回归【Regression】的,但它也可被用于其他的多元分析【Multivariate Analysis】(参阅 Frederick, 1999)。

### 启发性例子

为了使讨论更具体一些,我们假设一个涉及单个测量结果变量($Y$)和两个预测变量的($X_1$ 和 $X_2$)的回归问题。与这一例子关联的结果展现在了图 1 的维恩图(Venn diagram)中。

在这个例子中,三个量度的变量每个都有一个 400 的平方和【Sum of Squares, SOS】,对应于为每一个被画成面积为 20 × 20 个单位的量度的变量的方框。先来看 $X_1$ 和 $Y$ 的重叠部分,因为 $SOS_Y = 400$ 中的 100 个单位被 $X_1$ 预测了,所以 $r^2_{YX_1} = 100/400 = 0.250\,0 = 25.00\%$。$X_2$ 解释了 $Y$ 的 400 个单位中的 225 个,因此 $r^2_{YX_2} = 225/400 = 0.562\,5 = 56.25\%$。

然而多重相关【Multiple Correlation】($R^2_{[YX_1X_2]}$)的平方并不等于 25.00% 加 56.25%或 81.25%,因为这些预测变量是相关的($r^2_{X_1X_2} = 18.75\%$)。因此,$X_1$ 和 $X_2$ 的解释力中某些部分是二者共同的,故而在计算 $R^2_{[YX_1X_2]}$ 时不应该被“双重计数”。对于这一例子而言,$X_1$ 和 $X_2$ 一起解释了 $Y$ 分数中的变差的 $R^2_{[YX_1X_2]} = 68.75\%$(而不是 81.25%)。

但是这 68.75% 的效应量【Effect Size】是从何而来的呢? 下面的问题必须弄清楚:

1.观察到的效应量中属于 $X_1$ 的独特的解释力的作用有多大?

2.观察到的效应量中属于 $X_2$ 的独特的解释力的作用有多大?

3.观察到的效应量中属于 $X_1$ 和 $X_2$ 的共同的解释力的作用有多大?

这些问题正是共同性分析要弄清楚的问题。

在进行一个四步系列分析时,第一步,用预测变量的所有可能组合计算 $R^2$。在这个简单的例子中,只有三种可能的组合,$R^2_{[YX_1X_2]} = 68.75\%$,$R^2_{[YX_1]} = 25.00\%$ 和 $R^2_{[YX_2]} = 56.25\%$。

第二步,将在第一步求得的值插入在各种出版物中提供的共同性分析的公式(Rowell, 1996)。对于这一例子,$X_1$ 的独特的预测贡献($U_1$)应像下面这样计算:

$$U_1 = R^2_{[YX_1X_2]} - R^2_{[YX_2]}$$
$$68.75\% - 56.25\% = 12.50\%$$

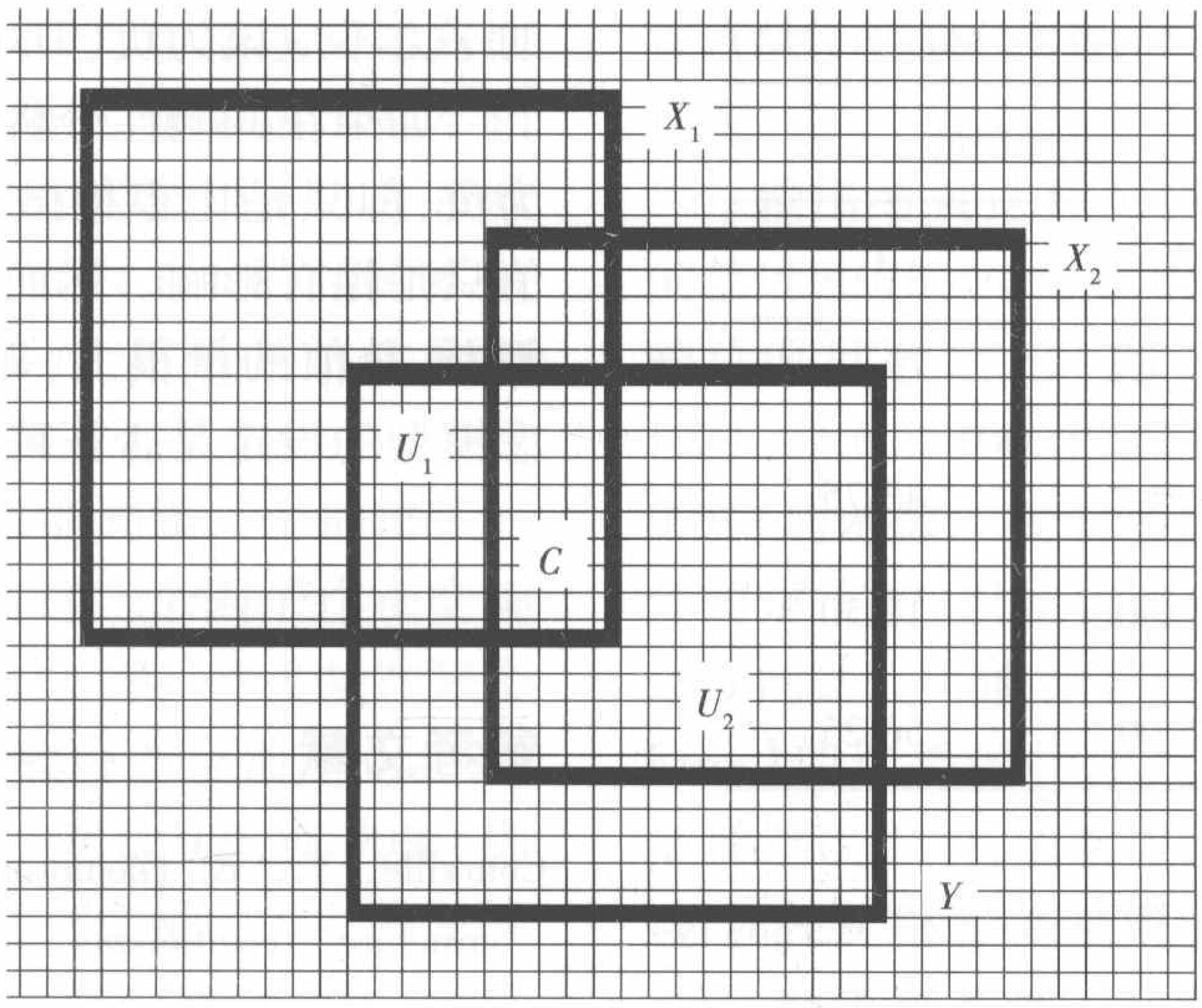



图 1 双预测变量回归共同性分析的维恩图

$X_2$ 的独特的预测贡献($U_2$)则是

$$U_2 = R^2_{[YX_1X_2]} - R^2_{[YX_1]}$$
$$68.75\% - 25.00\% = 43.75\%$$

$X_1$ 和 $X_2$ 的共同贡献($C_{12}$)是

$$C_{12} = R^2_{[YX_1X_2]} - U_1 - U_2$$
$$68.75\% - 12.50\% - 43.75\% = 12.50\%$$

共同性分析的计算公式以指数方式变得更为复杂，因为它有更多的预测变量(Rowell, 1996)。例如，在有三个预测变量时，公式就变成了

$$U_1 = R^2_{[YX_1X_2X_3]} - R^2_{[YX_2X_3]}$$
$$U_2 = R^2_{[YX_1X_2X_3]} - R^2_{[YX_1X_3]}$$
$$U_3 = R^2_{[YX_1X_2X_3]} - R^2_{[YX_1X_2]}$$
$$C_{12} = R^2_{[YX_1X_3]} + R^2_{[YX_2X_3]} - R^2_{[YX_3]} - R^2_{[YX_1X_2X_3]}$$
$$C_{13} = R^2_{[YX_1X_2]} + R^2_{[YX_2X_3]} - R^2_{[YX_2]} - R^2_{[YX_1X_2X_3]}$$
$$C_{23} = R^2_{[YX_1X_2]} + R^2_{[YX_1X_3]} - R^2_{[YX_1]} - R^2_{[YX_1X_2X_3]}$$
$$C_{123} = R^2_{[YX_1X_2X_3]} + R^2_{[YX_1]} + R^2_{[YX_2]} + R^2_{[YX_3]} - R^2_{[YX_1X_2]} + R^2_{[YX_1X_3]} - R^2_{[YX_2X_3]}$$

第三步，作为对计算的一种检查，$R^2_{[YX_1X_2]}$ 再一次被计算，计算方法如下所示：

$$R^2_{[YX_1X_2]} = U_1 + U_2 + C_{12}$$
$$12.50\% + 43.75\% + 12.50\% = 68.75\%$$

注意，正如图 1 所阐明的那样，$U_1 = 12.50\%$和 $C_{12} = 12.50\%$并非同一区域。

第四步，方差的分解随后被组织成一张分别阐明每一预测变量的独特和共同的预测贡献的表。表 1 列出了这一例子的有关结果。

## 常见的错误观念

人们在使用回归时一个常见的错误观念是预测变量应该是不相关的(即没有共线性【Collinear】或多重共线性【Multicollinear】)。在预测变量相关的时候，我们确实不能只用标准化的权数("beta 权数")导出有关探测到的效应量【Effect Sizes】的原点的合理解释。在预测变量相关时，标准化权数的标准误差会变大，这一点也是事实。

但是预测变量之间的相关【Correlations】并不一定就是问题，只要解释的策略使用得当。在现实中，实际上许多变量都是相关的，这时我们可考虑使用修正的预测变量。有一种这样的解释策略，它同时对标准化的权数和结构系数进行考察(Courville & Thompson, 2001)。

**表 1 平方双变量相关的方差分解(百分比)**

| 方差分解 | 预测变量 | |
|---|---|---|
| | $X_1$ | $X_2$ |
| $U_1$ | 12.50 | |
| $U_2$ | | 43.75 |
| $C_{12}$ | 12.50 | 12.50 |
| Sum $= r^2$ | 25.00 | 56.25 |

共同性分析是在预测变量相关的时候理解预测的动力学的又一种备择策略。注意,在预测变量完全不相关的时候,进行共同性分析将是不明智的,因为根据定义,这时每一个公共方差分解(即 $C_{12}, C_{23}, C_{123}$)都是零,而每一个独特的贡献(如 $U_1$)则等于一个相应的平方双变量相关(如 $r^2_{YX_1}$)。

## 负方差成分

在任何分析中,如果有过多的模型设定误差,结果就会受到影响。在共同性分析中,如果使用了错误的变量或错误的分析,情况也同样如此,结果同样也会受到影响。在某个方差成分为负的时候,就表明这样的情形可能已经发生,因为被解释方差统计值的平方不应该为负(Thompson, 1985)。

如果在共同性分析中,有几个方差成分为负,但只是比零稍微小一点,那么这个负值只是稍有影响。然而,如果根据研究者的判断,负值的量很大,那么就应该认真地考虑模型的设定是否正确。

——Bruce Thompson

(沈崇麟译校)

## 参考文献


Courville, T., & Thompson, B. (2001). Use of structure coefficients in published multiple regression articles: $\beta$ is not enough. *Educational and Psychological Measurement*, *61*, 229-248.

Frederick, B. N. (1999). Partitioning variance in the multivariate case: A step-by-step guide to canonical commonality analysis. In B. Thompson (Ed.), *Advances in social science methodology* (Vol. 5, pp. 305-318). Stamford, CT: JAI.

Rowell, R. K. (1996). Partitioning predicted variance into constituent parts: How to conduct commonality analysis. In B. Thompson (Ed.), *Advances in social science methodology* (Vol. 4, pp. 33-44). Greenwich, CT: JAI.

Thompson, B. (1985). Alternate methods for analyzing data from experiments. *Journal of Experimental Education*, *54*, 50-55.


# 公因子方差(Communality)

公因子方差是一个平方统计量,反过来,公因子方差可概念化为一个测量/观察的变量的方差,它在描绘模型中的潜/复合变量的时候有多大用处。一般用作公因子方差系数的符号是 $h^2$。

公因子方差通常都用于因子分析【Factor Analysis】,包括探索性因子分析【Exploratory Factor Analysis, EFA】和验证性因子分析【Confirmatory Factor Analysis, CFA】(Graham, Guthrie, & Thompson, 2003; Thompson, 1997)。不过公因子方差系数也在其他的多变量技术,如典型相关分析【Canonical Correlation Analysis】中计算(Thompson, 2000)。在变量分析的时候,将会为每一个变量计算单独的公因子方差系数。

公因子方差系数既可在分析完成之后

追溯估计，也可在预期分析中输入初始数据中分析。在追溯性应用中，如果因子是不相关的，那么一个量度变量的公因子方差系数则由计算一变量在所有因子上的结构系数（因子负荷）的平方得到，然后再将这些跨因子的平方结构系数加总。

例如，10 个变量可能同属一个抽出了三个因子的 EFA，然后对这三个因子作正交旋转【Orthogonal Rotation】。第一个变量在三个因子上的平方结构系数可能分别是 0.4，0.3 和 0.2。这一变量的 $h^2$ 就是 0.9，或 90%（即 0.4+0.3+0.2=0.9）。这一变量的方差的 90% 可以为这三个因子所复制。相反，这一量度变量的方差的 90%在将这三个因子作为一个组来描述时是有用的。

公因子方差系数也可被概念化为一个给定变量的最小信度系数【Reliability Coefficient】的下限估计量。如果一个量度变量有一个 $h^2=0.9$，那么不论这个变量的分数的信度是什么，信度都至少是 0.9。

一个分析的输入项目会影响输出的共通性。我们常常要作相关矩阵分析。不论最初置于矩阵对角线的项目是什么，它们总是在输入公因子方差的估计值。此后，因子才被抽取。在主成分分析【Principal Components Analysis】中，公因子方差系数是在这时计算的，而后分析才完成。

然而，在诸如这样的主轴因子分析中，这些公因子方差系数随后要被回代到相关矩阵的对角线，并再一次抽取因子。这一过程连续不断地重复，直至公因子系数逐步稳定。

输出结果迭代的影响是（a）变量分数的信度和（b）对角线上的项目数与矩阵中的总项目数之比（例如，10 个变量，就是 10/100=0.1；100 个变量，就是 100/10 000=0.01）的函数。我们还应该注意，我们要估计的是信度，以便将测量误差考虑在内，但过度的迭代可能会导致抽样误差【Sampling Error】使结果的扭曲过大。因此有些方法学者建议不用，或有限地使用迭代，以平衡这两种类型的误差的影响。

——Bruce Thompson

（沈崇麟译校）

＊也可参见因子分析【Factor Analysis】。

## 参考文献

Graham, J. M., Guthrie, A. C., & Thompson, B. (2003). Consequences of not interpreting structure coefficients in published CFA research: A reminder. *Structural Equation Modeling*, *10*, 142-153.

Thompson, B. (1997). The importance of structure coefficients in structural equation modeling confirmatory factor analysis. *Educational and Psychological Measurement*, *57*, 5-19.

Thompson, B. (2000). Canonical correlation analysis. In L. Grimm & P. Yarnold (Eds.), *Reading and understanding more multivariate statistics* (pp. 285-316). Washington, DC: American Psychological Association.

# 比较法（Comparative Method）

比较法从根本上说是针对个案或小样本 $N$ 的技术。如果是研究者对一项调查中的每个个案都有相当多的了解，而个案数量相对较少，就是比较典型的适用比较法的时候，而且通常也不得不这样做。掌握比较法基本特点的最好方法就是根据教科书介绍的社会科学方法论对它的差异加以考察。按教科书讲的来说：（a）社会研究的直接目

标是记录证明适用于大群体观察的一般模式;(b)个案和群体一般被视为给定的;(c)只要有可能,就鼓励研究者增加个案的数量;(d)常常假定研究者开始研究的时候就有界定清晰的理论和表述完善的假设;(e)建议调查者要将关注重点放在那些展现出有益于研究的变差范围的变量【Variables】上;(f)指示研究者要对竞争的自变量的相对重要性作出评价,以理解因果关系。

## 直接目标

教科书讲述的社会研究,通常都集中在记述那些描述观察的大总体特性的一般模式这一目标上。如果研究者可以证明一个或多个自变量与某个因变量【Dependent Variable】的关系,那么他们就能根据其所掌握的预测变量的分数来更好地预测个案在因变量上的值。通常,一般模式的研究采用的都是一个来自一个大总体的观察样本。研究者根据他/她对样本的分析来对较大的总体进行推论。

相比之下,比较研究【Comparative Research】关注的不是变量之间的关系或是推论和预测的问题,而是要弄清楚那些因为在某些方面实质上或理论上很重要而被选择的少数几个个案(Eckstein, 1975)。例如,一个研究者可能会用比较法来深入研究几次游击运动。假设这些运动都是被视为特别得到民众支持的成功的运动,为了弄清楚这些运动是如何成功的,研究者要对这些运动进行深入的研究。这一面向个案的比较研究和教科书的研究大相径庭。后者中的研究者要从这类游击运动的总体(假设可以确定这个总体)中去抽样,接着再用通用变量(generic variable)以及它们之间的关系来描绘这些运动(例如,外国经济支援的多少与民众支持程度之间的相关关系)的特征。

正如上述例子所示,最关键的差别就在于研究者的直接目标:研究者究竟是要力图理解具体的个案,还是记录描述某个总体的一般模式?这个差别源于一个源远流长的分野,存在于科学的所有领域,而不只是在社会科学中。格奥尔格·亨利克·冯·怀特(von Wright, 1971, pp. 1-33)在《解释与理解》(*Explanation and Understanding*)一书中指出,在观念史上,那些使一个解释在科学上得到认同所必须满足的条件存在两种传统:一种传统他称为"终极目的论的传统",专注于理解事实的问题;另一种传统叫作"因果机械论传统",专注于预测的问题。比较研究和"教科书式"社会研究的不同目标与这一基本分野颇为相似。

## 个案和总体的构成

教科书介绍的社会科学方法论几乎不会去考察个案和总体的构建问题。一个平常的个案就是被调查者个人;一个平常的总体由地域、时间和人口的界限来界定(例如,2003年美国的成年人)。教科书要解决的问题的关键是,如何从一个正在观察的很大的总体中导出一个有代表性的、假定为研究者处置的样本。

相比之下,采用比较法的研究者把个案看成事件与结构(例如游击运动)之富有意义的复杂构成,他们的个案一般来说都是宏观层次的。而且,他们把个案当成特意定义和挑选出来的个别的、完整的存在,而非来自大量貌似相同选择的、随机抽出的同质性的观察数据。许多比较研究从一个看似简单的想法入手,即相似背景下的社会现象(例如组织、国家、地区、文化等)可能彼此足够相似,让我们得以对它们进行比较和对照。在这个论述里,"可能彼此足够相似"这个句子是个非常重要的组成部分。采用

比较法的研究者在开始调查时对相关个案的说明实际上只是一个工作假设而已，也即最初选择的那些个案实际上都是同一事物（例如游击运动）的不同实例且足够相似，让我们得以进行比较。在研究的过程中，调查者可能会再作别的决定，把一些个案或者一整类个案都剔除掉，因为它们看来并不属于核心个案那一类。有时候，这个筛选个案的过程会造成相关个案的规模扩大，指导概念的范围也会相应地扩展（例如从游击运动扩展为激进的反对运动），也可能会发展出一套处理各种个案的类型学（例如游击运动的各种类型）。

一般来说，这种对个案的筛选和分类是结合概念的构造和阐释来进行的（Walton，1992）。随着相关个案集的边界慢慢调整和明晰，概念也得到了修正和细化（Ragin，1994）。在理念和证据的对话中常常会出现重要的理论区别。当调查者对他/她研究的现象了解得更多，细化了研究的指导型概念和分析策略之后，研究者对"我的个案是什么"以及"这些个案属于什么"这两个问题的回答在研究过程中可能会发生改变。

## 个案的数量

每个定量的社会研究课程都有非常关键的一课，就是"个案越多越好"。越多越好主要包括两个方面：首先，研究者就是为了采用定量研究，也必须满足个案数量的下限，一般人们所说的 $N$ 是 30～50 个个案。其次，个案 $N$ 数量越少，数据就越得符合统计方法的假定，比如说，变量正态分布假定或子群方差基本相同假定。但小样本 $N$ 几乎都无法满足这样的假定，在个案都是宏观层次的时候，情况尤其如此。教科书这种对大样本 $N$ 的偏爱符合一个隐含的假定，即所有的个案都是经验上给定的，而非研究者构建的，而且个案多的是。就此而言，唯一的问题就是研究者是否愿意、是否能够搜集尽可能多的数据，如果不是成千上万，至少也得数以百计。

相比之下，比较研究常常是以其对现象的关注为定义，这些现象是有趣的，因为很少见，也就是说，恰恰是因为个案数量很少。一般来说，这些现象规模巨大，并受历史的限定，在任何意义上都不是普通的。受历史记录所限，与某项调查相关的这类个案的总体通常不过区区几个。它和教科书式的社会科学的重大差异源自一个简单的事实，也即许多社会科学家和他们的阅听人感兴趣的现象都是历史和文化上的重要现象。去争论社会科学应该只研究普遍、大量存在的个案，或者可以独立于历史文化环境的个案会大大限制社会科学的领域和价值。

## 理论检验

按照惯例，介绍社会科学方法论时会着重强调理论检验。实际上，其理论检验的定位经常被说成社会科学之所以是科学的原因。它建议研究者按照科学的方法，独立于经验证据的分析来发展他们的假设。它进一步假定现有的理论都构建（formulated）得足够好，可以明确地得到可检验的假设，而社会科学知识主要是通过拒绝那些总是无法得到经验支持的理论而推进的。当然，这种教科书的观点并没有限制那些更多的以归纳的方式产生的观点出现在社会科学中，但这种观点在得到明确的检验之前都会受到怀疑，而检验这种观点所用的数据必须不同于生成该观点的数据。

毋庸置疑，理论在社会研究中处于核心位置，实际上，几乎所有的社会研究都在某个方面非常依赖理论。不过，一般不可能把理论检验的范式应用于小样本 $N$ 的、以个案为目标的研究。多数比较研究都带有很大的归纳成分。研究者通常会关注少数几

个个案,可能会花很多年来研究它们,延伸其知识的深度。在这个研究个案的过程中会产生无数的想法和假设,甚至在研究的过程中对总体的定义本身(这些个案属于什么?)都可能会多次改变。而且,因为比较研究中相关个案的数量通常都很少且受制于历史记录,确实不可能用个案中的"另一个例子"来检验源自深度研究的观点。常常是研究一开始所包含的个案就是全部的相关个案。

大多数比较研究的直接目标,就是对重要的历史和文化现象的"方式"(how)作出解释。例如,游击运动是如何形成的?理论给经验研究提供重要线索和指导性概念,起到了重要的确定方向的作用,但现有理论几乎构建得都不够好,无法在比较研究中提供明确的假设。比较研究的主要科学目标不是理论检验本身,而是概念的构建、阐述和细化。比较研究中主要的工作常常都集中于构建研究的个案,并进一步细化适用于所选个案的概念(Ragin & Becker, 1992)。

## 因变量

教科书介绍社会研究时有一个最基本的概念就是变量——让一个个案和相邻个案不同的某个特性或某个方面——以及与之相关的寻找不同个案间变量的关联模式的想法。比如说,是否较为富裕的国家所经历的政治动乱比贫穷国家要少?如果是这样,那么社会科学家可能会提出,不同国家间政治动乱的情况(因变量或者结果变量)可以部分地用国家财富的变化(自变量或原因变量)来解释。这些有关变量的观点隐含了一个原则:社会科学家希望解释的现象在他们所研究的各个个案中必定会变化;否则就没什么可解释的。由此可进而得出结论,如果各个研究个案的结果没什么不同,那么研究者几乎就不可能发现导致这些结果的原因。

但是,采用比较法的研究者经常特意去选择那些结果没有差异的个案。例如,一个研究者可能会尝试解释游击运动是如何形成的,对几个成功的实例进行研究。从教科书的社会研究的观点来看,这位研究者却犯了一个重大的错误——选择了一些因变量差异很小的个案,假如还算得上有差异的话。毕竟所有的个案都是成功形成了游击运动的实例。这个研究缺少可变的因变量,似乎根本就不存在分析的可能。

对于研究"同一个事物"各个实例的这些担忧理由很充分。不过,这都是源于对比较研究的逻辑的错误理解。采用比较法的研究者不是用一个变量的变化来解释另一个变量的变化,而通常是在他们的各个个案中——也就是在结果的各个实例中——去寻找共同的因果条件(Robinson, 1951)。比如说,假如所有成功形成的游击运动都出现在土地分配高度不平等、严重依赖农业出口的国家,那么这些结构因素可能对解释游击运动是非常关键的。实际上,在对成功形成的游击运动的比较研究中,在构建了类别、指定了具体相关个案之后,第一步就是寻找各个实例中共同的因果条件。尽管用常量来解释常量(例如寻找相似实例都有的共同原因)在比较研究中很常见,对关注变量间关系的技术来说它却还很陌生——对后者而言,个案与个案之间的因果条件和结果存在很大差异。

有时候比较研究的研究者可能会找到可以与其正面个案比较的反面个案。比如,在对成功开展的游击运动的研究中,早期就失败了的游击运动就可以作为反面个案。如果研究的目标是要找到产生某一结果的充分条件,反面个案就是必要的(Ragin, 2000)。比如说,如果进行比较的研究者提出,对于成功开展游击运动而言,某种具体

的因果条件组合是充分条件,那么说明那些功亏一篑的游击运动中(即在相关的反面个案中)没有满足这些条件就很重要。不过要注意,只有当研究者能组建一个相关反面个案集合的时候,才可以评价一组因果条件是否充分。要这样做通常会很困难,因为经验范畴要么太宽泛,要么太模糊(例如"缺乏成功的游击运动"这个类别),要么无法找到或收集到证据(例如有关中途失败的运动的数据)。如果这些障碍都可以克服,那么我们还必须明白,正面个案的构建和分析是构建和分析反面个案的必要的前提条件(Griffin et al., 1997)。因此,即使进行比较研究的研究者计划考察各个个案不同的结果(即在一些个案里有,而在另一些个案里却没有),常常也需要先研究所有具有这个结果的个案(即正面个案)。

## 因果分析

按照教科书对社会研究的逻辑的讲法,大多数社会数据分析的主要目标就是评估自变量作为导致因变量变化的相对重要性。例如,某研究者可能想要了解民主制度的设计和被认知到的合法性究竟哪个对民主制度的持续存在影响更大。按照这种观点,各个原因变量相互竞争,都对某个结果变量的变化进行解释。在这场竞争中,一个有力的竞争者就是与因变量存在强相关而与其他竞争自变量只是微弱相关的自变量。

相比而言,比较研究的研究者通常只从组合的角度考察各种原因:在每个个案中,相关的各个原因是如何结合并生成我们讨论的结果的?比较研究的研究者不把各个原因视为竞争性的因素,而是将其当作原材料,它们相互结合产生了所研究的质性的结果。任何特殊的原因条件都取决于其他条件存在或不存在,而几种不同的条件可能满足一种综合的因果要求——也就是说,在一个更为抽象的层级上,两个及两个以上不同的原因可能是等价的。构建和选择好某一结果的相关实例之后(例如成功开展的游击运动),如果可能,也定义清楚相关的反面个案,比较研究的研究者的任务就是处理每个实例背后的因果力量,并对个案之间的异同点予以特别关注。每个个案的细节都在理论概念、实务知识和兴趣的指导下一一加以考察,以回答在每个正面个案中结果"如何"出现,在每个反面个案中为何不出现(假设可以确信无疑地找到这些个案)。比较研究的一种常见结果就是各种原因的不同组合可能会产生相同的结果(Ragin, 1987; Mackie, 1974)。

## 总结

从构建个案到考察与一组个案的结果相关联的不同条件的组合,本词条简要勾勒的对比粗线条地展现了比较研究的特点。它的关键点就在于比较研究不太符合教科书上有关社会研究的正确做法的教诲。社会科学家和他们的阅听人通常感兴趣的是理解在某种意义上具有文化和历史重要性的社会现象。这些现象常常都很少见,因而只能通过面向个案的深度研究才能理解。

——Charles C. Ragin

(杨可译　沈崇麟校)

*也可参见个案研究【Case Study】。

## 参考文献

Eckstein, H. (1975). Case study and theory in political science. In F. I. Greenstein & N. W. Polsby (Eds.), *Handbook of political science: Vol. 7. Strategies of inquiry* (pp. 79-137). Reading, MA: Addison-Wesley.

Griffin, L., Caplinger, C., Lively, K., Malcom, N. L., McDaniel, D., & Nelsen, C. (1997). Comparative-historical analysis and scientific infe rence:

Disfranchisement in the U.S. South as a test case. *Historical Methods*, *30*, 13-27.

Mackie, J. L. (1974). *The cement of the universe: A study of causation*. Oxford, UK: Oxford University Press.

Ragin, C. C. (1987). *The comparative method: Moving beyond qualitative and quantitative strategies*. Berkeley: University of California Press.

Ragin, C. C. (1994). *Constructing social research*. Thousand Oaks, CA: Pine Forge Press.

Ragin, C. C. (2000). *Fuzzy-set social science*. Chicago: University of Chicago Press.

Ragin, C. C., & Becker, H. S. (1992). *What is a case? Exploring the foundations of social inquiry*. New York: Cambridge University Press.

Robinson, W. S. (1951). The logical structure of analytic induction. *American Sociological Review*, *16*, 812-818.

von Wright, G. H. (1971). *Explanation and understanding*. Ithaca, NY: Cornell University Press.

Walton, J. (1992). Making the theoretical case. In C. C. Ragin & H. S. Becker (Eds.), *What is a case? Exploring the foundations of social inquiry* (pp. 121-137). New York: Cambridge University Press.

# 比较研究(Comparative Research)

比较研究的主要目标是确定社会实体之间的异同点。

比较研究追求的是对不同国家、文化、社会和制度进行比较。学者们在使用这个术语时,其意义有所不同:对有些人来说,比较研究严格限定为对两个或多个国家进行比较(也被称作"跨国研究"),但其他学者倾向于扩大比较的范围,把许多不同种类的社会与/或文化实体的比较也包括进来。而还有一些学者在使用比较这个术语的时候,把民族国家或其他文化与社会边界之内或之间的亚文化以及其他社会实体的比较也包括了进来。

尽管学者们对比较的定义远未达成共识,但其趋势还是倾向于把社会科学中的比较研究定义为系统比较两个或多个社会、文化或者国家的研究。在实际操作中,国家间的比较占据主导。比较研究的目的在于更高的普遍性:比较研究的目标就是以被比较的社会实体的异同点,尤其是其思维和行动方式的特征,其特有的态度、价值观和意识形态以及它们的社会结构的内部要素的异同点为根据,发展概念和一般法则,在此之后,这就可作为一种我们更好地理解和感知其他社会实体的工具。

## 历史发展

社会学的奠基人都是进行比较研究的学者。卡尔·马克思、马克斯·韦伯、埃米尔·涂尔干以及阿历克西·托克维尔,虽然仅列举几例,但却无一例外,都是坚定不移的比较法使用者,无论他们研究的是角色、制度、社会、国家、文化、群体还是组织。以涂尔干为例,他希望找到一个普遍的"法则"来解释不同国家和职业之间自杀率的差异。不过,比较研究在20世纪初曾一度衰落,因为有段时间学者们坚持认为,社会研究在方法论上对精度的追求应该越来越高,而比较研究的方法论资源(methodological resources)在他们看来似乎不够严谨。但自第二次世界大战以来,一方面由于比较研究在方法论和方法论工具上得到改进;另一方面受第二次世界大战后的国际环境影响,比较研究又在社会研究中取得了核心位置。那种国际环境(例如曾经两极分化的共产主

义阵营和非共产主义阵营)和我们后来的国际化,都对社会科学比较研究的重要回归作出了贡献。

## 应用

比较研究的目标有很多,但一个重要的任务就是支持理论建构,并对其作出贡献。尽管是理论框架驱动着比较研究的建设工作,但这类研究的结果经常会推动理论的革新。它的另一个重要任务是支持政策制订,主要是就国家层面而言。国家 B 是否可以用国家 A 的战略方针来处理一个既定的社会问题(例如滥用毒品)？或者在这两个国家之间存在着特别的差异,使得这样的战略方针不可行？比较研究还有一个目标是弄清某个既定概念(例如宗教信仰)的同一维度是否可以当作通用的社会指标来使用。某个已知概念是否概括了所有国家(或者其他社会实体)的情况？

所有比较研究都关乎视角。一个国家的研究者需要一种采纳和理解其他国家的同行观点的工具。随着世界日益全球化的进程,对这种理解的需要应该是很清楚的:国家政策需要考虑各种受到政策实施影响的全球合作伙伴,并把他们的需求涵括进来。政策制订者是否具有赖以确保合作和承诺的必要视角？

社会科学比较研究在各种方法论指导下进行,从严格的定性研究到苛刻的定量研究。个案研究以及相似类型的分析代表了描述性的定性研究方法,其研究者可以是被研究的社会实体的成员,也可以是在这个被研究的社会实体之外的研究者。近年来,比较研究中的定量研究占据了相对主导的地位,因为可以有工具(特别是计算机、统计和数学工具)来处理严格的比较研究在方法论上的许多复杂之处。

实证研究的调查越来越普遍,尽管通常很复杂而且耗资不菲。复杂性还包括语言障碍在内,这使得准备可供比较的问卷很困难,还有,“文化”障碍也常常使得准备可供比较的问卷几乎不可能:如果你问美国人“你信上帝吗?”几乎每个人都会说信,而如果你在日本问相同的问题,大概只有 1/3 的人会说信,这意味着对两个国家而言,相同的仅是有着一致回答的比例,此外,在两个环境中问这个问题几乎没有比较的意义,因为从中推论不出什么有意义的结果。同样地,在一个不以市场为区别基础的社会群体中询问就业和失业也是文不对题。就这个例子来说,两个社会群体对“就业”这个概念不存在“概念等价性”。

近年来,统计方法的进展极大地提高了社会科学研究者分析原始数据和/或重新分析二手数据的能力。比如说,诸如同期群分析【Cohort Analysis】这样的方法发展了出来,可以降低严格的样本选择的需要,大大便利了对集体数据异同点的探查和说明。当然,和所有严格的科学研究一样,社会科学比较研究也必须满足统计的效度和信度等一般要求。

正在进行的比较研究调查的例子包括 1945 年以来的盖洛普民意调查(the Gallup Polls, since 1945), 1972 年以来的社会综合调查(the General Social Survey, since 1972), 1972 年以来的欧盟调查局调查(the Eurobarometre, since 1973), 1994 年以来欧洲社区家庭调查(the European Community Household Study, since 1994), 以及始于 1984 年的国际社会调查项目(the International Social Survey Program, ISSP),它已经在超过 37 个国家作了一般社会态度调查。日本国民性研究(the Japanese National Character Study)从 1953 年开始,每 5 年做一次调查。所有这些调查都会定期进行,而且常常在多个国家进行。

为什么各个国家在有的特征参数上不

同,在别的特征参数上又一样?这个相同的问题也可以拿来问一些其他组的社会文化群体,既可以跨越地区,也可以在国家或种族边界内。因为国界不一定总是与种族边界相同,比较研究也不会总是限于简单的国界。这个现象与近来出现的东欧国家特别有关。

比较研究的终极目标就是通过对理论提出越来越严格的要求,坚持研究必须可在不同的社会群体间重复,促进普遍准则的跨社会群体应用,对独特的模式作出解释,从而为理论建设或重建作贡献。

### 未来的方向

如果按照全球化所预示的那样,全球在社会和文化上将最终趋同,比较研究也会再次衰落,因为没有几个独立的社会实体可供比较。不过,与此同时,在世界上还有很多变化和可变性,而比较研究则有望进一步对此向我们作出解释,并进一步增进我们彼此之间的理解和认知。

——Masamichi Sasaki

(杨可译　沈崇麟校)

### 参考文献

Hantrais, L., & Mangen, S. (Eds.). (1996). *Cross-national research methods in the social sciences.* London: Pinter.

Inkeles, A., & Sasaki, M. (Eds.). (1996). *Comparing nations and cultures.* Englewood Cliffs, NJ: Prentice Hall.

Przeworski, A., & Teune, H. (1970). *The logic of comparative social inquiry.* New York: John Wiley.

Smelser, N. (1997). *Problems of sociology.* Berkeley: University of California Press.

## 参照组(Comparison Group)

参照组指的是在测试某种处理效果的实验里,没有经过处理【Treatment】或者经过其他处理的一组单位(例如人、教室)。参照组是作为一种反事实【Counterfactual】的因果推论资源来使用的。一个反事实就是一个没有发生的事(与事实相反);就这里而言,经过处理的反事实就是没有经过处理。根据反事实因果推论的理论,处理的效果就是经过处理的那些人的结果,与那些在同一时间未经过处理的同样的人群的结果之间的差异。遗憾的是,不可能观测到反事实,也就是说,不可能同时对同样的人既进行处理又不进行处理。于是,研究者通常会采用参照组作为真正的但无法观测的反事实的替代,然后他们就可以比较处理组的结果与参照组的结果处理效果的差别。为了把这个做好,研究者应该保证参照组的所有单位都尽可能地类似于处理组的单位。把各个单位随机分配【Randomly Assignment】使处理组和参照组可以提高两组所含的单位的相似性,退而求其次的话,也有许多种准实验【Quasi-Experimental】法可以提高相似性,例如根据被认为与结果相关的变量来对各个单位进行匹配【Matching】(Shadish, Cook, & Campbell, 2002)。

### 参照组的种类

一个处理组可以与许多不同种类的参照组进行比较,例如,(a)无处理组,有时候也称为控制组【Control Group】,研究对象没有受到任何介入,没有改变其环境;(b)常规处理组,即提供的是标准照料或处理;(c)安慰剂【Placebo】组,研究者模仿正常处理的方式进行介入,但并不包含处理的活性

成分;(d)候补名单组,承诺研究对象他们会在处理组实验结束后得到同样的介入处理;以及(e)替代处理组,研究对象受到一种不同于实验主要关注的处理的替代性介入(处理)。就最后一种情况来说,替代处理常常是目前解决某种问题的最好的处理方式——"黄金标准"——以便了解正在实验的处理方法是否可以改进其结果。另一区别在于,组间【Between Groups】比较中的参照组是不同于处理组的一群研究对象,但在被试者内【Within-Subjects】设计比较中,同一批研究对象会先后作为参照组和处理组。被试者内比较需要较少的单位来测试某一处理,但可能会遭遇到各种处理污染效应(treatment contamination effects),例如练习效应(practice effect)、疲劳效应,以及从一种条件到另一种条件的处理延滞效应【Carryover Effects】。

## 历史发展

可能早在远古时代,参照组就被用来协助因果推论。罗森塔尔和罗斯诺(Rosenthal & Rosnow,1991)讲了一个古埃及的地方法官的故事,他平常总是用毒蛇来处决罪犯。这个地方法官在调查为什么所有的罪犯被蛇咬了之后都不会死,他发现有人给他们吃了圆佛手柑。为了查明让罪犯们免于一死的是不是圆佛手柑,他只给一半的囚犯吃了圆佛手柑,然后把所有囚犯都拉去让蛇咬。在这个例子里,没有吃圆佛手柑的研究对象就是参照组,他们在被蛇咬之后都死了,而在吃了圆佛手柑的处理组中一个人也没死。

## 例证

安东尼奥等人(Antoni et al.,1991)采用了一个不进行处理的参照组,来考察男同性恋者在被告知其感染艾滋病之前的认知行为压力管理处理对于他们的压力水平的影响。处理组的人接受了减轻压力的训练。然后把他们的压力水平与没有接受减压训练,即未经处理的控制组的压力水平进行比较。处理的效应便是在得知感染艾滋病之后处理组和参照组之间压力水平的差异。

莫尔等人(Mohr et al.,2000)采用了照常规处理的参照组来考察认知行为处理对减少多发性硬化(MS)症病人罹患抑郁症的效果。处理组接受了一种新的抑郁症的实验疗法,而参照组接受的则是罹患抑郁症的MS症患者的常规疗法。接受抑郁症的标准疗法者与接受抑郁症新疗法者之间的抑郁症得分的差异就是这里产生的处理效应。

基夫拉罕、马拉特、弗罗姆、科佩尔和威廉姆斯(Kivlahan, Marlatt, Fromme, Coppel, & Williams,1990)采用了一个候补名单的控制组来考察技能训练项目对减少大学本科生饮酒的效应。实验结束之前一直没有对参照组进行任何处理,实验结束后,参照组也得到了与处理组一样的处理。

——William R. Shadish
M. H. Clark
(杨可译 沈崇麟校)

*也可参见实验【Experiment】、处理【Treatment】。

## 参考文献

Antoni, M. H., Baggett, L., Ironson, G., LaPerriere, A., August, S., Kilmas, N., et al. (1991). Cognitive-behavioral stress management intervention buffers distress responses and immunologic changes following notification of HIV-1 seropositivity. *Journal of Consulting and Clinical Psychology*, *59*(6), 906-915.

Kivlahan, D. R., Marlatt, G. A., Fromme, K., Coppel, D. B., & Williams, E. (1990). Secondary prevention with college drinkers: Evaluation of an alcohol skills training program. *Journal of Consulting and Clinical Psychology*, 58 (6), 805-810.

Mohr, D. C., Likosky, W., Bertagnolli, A.,

Goodkin, D. E., Van der Wende, J., Dwyer, P., et al. (2000). Telephone administered cognitive-behavioral therapy for the treatment of depressive symptoms in multiple sclerosis. *Journal of Consulting and Clinical Psychology*, *68*(2), 356-361.

Rosenthal, R.,& Rosnow, R. L. (1991). *Essentials of behavioral research: Methods and data analysis* (2nd ed.). New York: McGraw-Hill.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton-Mifflin.

## 竞争风险(Competing Risks)

一组互斥离散的结果或事件叫作竞争风险。这一术语主要出现在存活率分析【Survival Analysis】和事件史分析【Event History Analysis】的文献中。

早在18世纪,对存活和它的对立问题(如死亡、故障)感兴趣的学者们就认识到,引起死亡和故障的原因是多种多样的。例如一个人可能由于接触性传染的疾病、器官疾病(如癌症)、意外事故、自杀或蓄意的暴力行为(如谋杀、战争)而死亡。这些备择的死亡原因阐明了死亡的竞争风险。令学者们常感兴趣的并非只是各种原因结合在一起的死亡率和故障,他们也对各种各样的原因的死亡、故障和与特定的原因关联的死亡或故障率感兴趣。这就是说,他们希望对某些事件,如死亡或故障的竞争风险作出区别。

虽然竞争风险这一理念的语言和应用最初始于医药卫生、人口统计学和工程等领域,以及与之相关的涉及死亡或故障的领域,但它们也被广泛地应用于社会科学研究中。例如,研究劳动力的市场进程的研究者,不仅希望分析总的离职情况,也想分析人们之所以离职,是否因为在组织内部的提职或降职、平级调动、退职、解雇、退休或死亡。研究企业成功之道的学者也许不仅想考察所有形式的企业“死亡”,同时也想考察企业停止了所有活动,被其他公司收购或与其他公司平等合并。研究国际冲突和战争的社会科学家也许不仅希望考察整个冲突的终止,同时也希望考察冲突是否通过和谈过程终止,还是通过占领或因僵持而终止。在每一种情况中,都希望能对各种竞争风险加以区别。

人们经常关心的一个问题是,改变一个或多个竞争风险(发生)的可能性引起的后果是什么。一个典型的问题是:制止或完全消除一种特定的风险(如死亡或故障的特定原因)的后果将会是什么?为了说明这一问题我们来举一个例子。公共卫生法和对接触传染病普遍免疫法的引进极大地降低了感染接触传染病的可能性。正因为这样,许多最为致命的传染病的死亡风险和死亡率大大减少。降低来自任何竞争风险的死亡率有利于延长寿命,导致平均寿命的增长。一般而言,降低一种竞争风险的危险率,将会延长所有原因组合导致的事件发生的平均时间。

事件的竞争风险经常与年龄和时间高度相关。正因如此,降低一种竞争风险的危险率,并不意味着另一种竞争风险的危险率的降低,二者可能有很大的不同。竞争风险率的危险率取决于每个特定原因的危险率中的年龄和时间相依性的模式。例如,老年人心脏病和癌症发作的可能性大大高于儿童和青年,而婴儿和儿童较成年人更容易得传染病和发生意外事故。结果是,某种程度

的降低传染病的死亡风险,比同等降低的心脏病的死亡风险更能延长人的寿命。因此改进环境卫生和免疫系统,将使青年人死亡的总风险陡降,从而大大增加平均期望寿命(和延长处在工作年龄的成人的寿命)。相反,降低心脏病的死亡风险,则特别有利于延长老年人的余生,尤其是延长退休后的寿命。

降低传染病,特别是降低儿童容易感染传染病的死亡率的一个间接后果是,增加了心脏病、癌症和其他老年人容易患上的疾病的死亡的相对可能性。改进环境卫生和免疫系统确实不能消除其他死亡的竞争风险,它们毕竟对那些其他原因的死亡率没有直接的影响。一般来讲,直接抑制或消除一种竞争风险,会增加那种被压制或消除的竞争风险之外的那些竞争风险引起的死亡或故障发生的相对可能性。相反,增大一种竞争风险(如谋杀、战争)会降低其他竞争风险(如疾病)引起的死亡或故障的相对可能性。

分析竞争风险和测量的协变量在一个竞争风险而非另一个竞争风险上的发生和时机的效应模型主要有两种类型。这两种类型的根本差别在于,风险是否发生在离散的时点,还是任何连续的时间的瞬息。在第一种情况中,使用最为普遍的统计模型是多项 logistic 回归模型【Multinomal Logistic Regression Model】,或条件 logistic 模型(相关内容请查阅 Theil, 1969; McFadden, 1973;Maddala, 1983)。在第二种情况中,即所谓的事件史分析,是分析竞争风险的经验数据的标准方式。事件史分析中使用的特定模型包括各种类型的比例或不成比例的危险和转换率模型(用于事件史分析的主要备择模型的概述请参见 Tuma & Hannan, 1984)。

两种建模法的关键问题在于,竞争风险是否在统计上是独立的。我们用一个例子来阐明。一个患有严重疾病,如艾滋病的人特别容易因感染一种疾病死去。如果情况真的是这样,这两种死亡原因(如艾滋病和一种传染病)在统计上就不是独立的。相反,意外事故、谋杀和战争引起的死亡可能真的被认为是与慢性病,如艾滋病引起的死亡风险是独立的(或几乎是独立的)。

在使用离散多项 logistic 回归模型分析各种结果的特定原因的发生时,研究者可以检验不相干的备择选择的独立性(Debreu, 1960;Maddala,1983),也就是不同的竞争风险在统计上是否独立。如果竞争风险在统计上是独立的,那么多项 logistic 回归模型的估计值将恒等于(在抽样变异性范围内)一系列排除了一或多个竞争风险的较为简单的 logistic 回归模型的结果。有关这些检验的讨论请参见参考书目中的有关文献(Hausman & McFadden,1984)。

在事件史分析中,研究者可以估计与每一竞争风险相联系的即时的转换率(即特定原因的危险率)。然后可将这种分析得到的结果,在所有其他竞争风险(在统计上)独立的假定下,用于推论消除一个特定的风险的后果。同样,这些结果也可用于推论导致和某一风险相联系的转换率的上升或降低(即估计的转换率乘以某个假设的大于或小于 1 的分数)的后果。这样的推论在竞争风险(在统计上)独立的假定下是正当的(David, & Moeschberg, 1978)。齐阿蒂斯(Tsiatis, 1975)证明,即使使用全部和完整的竞争风险发生和择时的事件史数据,也不可能完全确认竞争风险在统计上是真的独立的。

——Nancy Brandon Tuma

(沈崇麟译校)

## 参考文献

David, H. A., & Moeschberg, M. L. (1978). *The theory of competing risks* (Griffin's Statistical Monograph # 39). New York: Macmillan.

Hausman, J., & McFadden, D. (1984). Specification tests for the multinomial logit model. *Econometrica*,

*52*, 1219-1240.

Maddala, G. S. (1983). *Limited-dependent and qualitative variables in econometrics*. Cambridge, UK: Cambridge University Press.

Tsiatis, A. (1975). A nonidentifiability aspect of the problem of competing risks. *Proceedings of the National Academy of Sciences*, *72*, 20-22.

Tuma, N. B., & Hannan, M. T. (1984). *Social dynamics: Models and methods*. Orlando, FL: Academic Press.

## 复杂数据组(Complex Data Sets)

所谓数据是复杂的有两种不同的含义，即统计上复杂的和结构上复杂的。我们可以通过将它们和一组全国性的简单数据作比较来对此作出解释。这种数据是由这样一组或对应于一个总体的等概率的随机样本(该样本中的每一实体都是独立地选取的);或对应于一个完整的总体的实体组成的。这个数据组包含每一实体的相同的完整的量度。例如,它可能是由一个个体的随机样本组成的。样本中每一个体都回答了一组有关他们自己的政治态度的问题。在数据组背离了这些假定时,数据组就被认为是复杂的。

在有关抽样的某些标准的假定不能成立时，如果没有作调节,那么标准的统计分析就可能误导结果,这时就会出现统计复杂性。斯金纳、霍尔特和史密斯(Skinner, Holt, & Smith,1989)对这种统计复杂性作了一个概述。如果在一个抽样调查中,回答人都聚集在一组有限的地区,这样的事常常会因为降低实地调查的费用而发生,那么由此便会导致对标准抽样假定的背离。这时观察便不再是独立的。这将意味着,用于计算显著性检验的标准误差,将需要对照以往使用的标准公式进行调整。例如英国社会态度调查,乍一看,它就是上面设想的那种简单数据组,而实际上它却有着一种聚类设计。空间和时间系列数据组则会缺乏另一种形式的独立性。在样本并非等概率地抽取的时候,则会发生另一种对标准抽样假定的背离。估计值便会因此而产生偏倚,因而需要进行调整,例如通过加权进行调整。发生在雇主调查中的这种背离是相当典型的。在雇主调查中,一大群雇主有着很高的被选择的概率。同样,如果由抽取的样本个体给出的回答的概率存在变化,无论是对整个访谈还是对单个问题,那么便可能会因此而产生偏倚,故而需要对此进行调整。

在一个单一的数据组中我们有着不同种类的实体的信息,结构上的复杂性便会因此产生。需要特别指出的是,许多数据都有着多重结构,以反映社会现实的某些重要方面。这样的例子不仅包含层次结构,如个体在家庭中的地位,工人在企业中的地位,以及儿童在学校中的地位;也包括同一个体在不同时点的纵向数据中的多重观察。在家庭预算或劳动力调查中,收集住户中所有人的信息便是一个十分典型的例子。也有些层级结构不是十分清楚的数据,例如社会网研究的数据。在每一主要实体有着不同的信息量的时候,数据组同样也可被视为一种复杂的数据组。例如在追踪调查中数目不同的观察,在生育史调查中,每个妇女不同的生育孩子数,或工作史调查中每个人从事过的不同工作数。例如诸如英国住户追踪调查(the British Household Panel Survey)这样的调查,因为人员损耗【Attrition】和儿童成长为合格的被调查人的缘故,每期调查的个体都会有所不同。

这种结构上的复杂性会造成多种后果。

首先,这意味着分析者要考虑清楚,自己的分析单位应该是什么。可以用这些结构,将较高层的单位作为分析单位处理,而将较低层次的概括性信息附加其中(例如家户层次的分析,使用户内个人的收入信息,或生育孩子数分析女性)。其次,给出的这些选择,数据可能需要作大量的重组,以便建立统计软件通常使用的矩形的数据组。最后,必须考虑结构的复杂性造成的统计后果。这可能只是必须对观察的非独立性进行调整,例如,更为积极的做法是用多层次建模法来利用结构的长处(Goldstein, 1995)。

——Nicholas H. Buck

(沈崇麟译校)

## 参考文献


Goldstein, H. (1995). *Multilevel statistical models*. London: Edward Arnold.

Skinner, C., Holt, D., & Smith, T. M. F. (1989). *Analysis of complex surveys*. Chichester, UK: Wiley.


# 成分分析(Componential Analysis)

成分分析(CA)假定任何一个给定的词汇的含义都由一组独特的有意义的特性予以最好代表。CA 法试图比较来自同义域的词汇的含义,以发现为区别它们的不同之处所需的最少几个特性。例如,在亲属域内,一个人的母亲和父亲、姐妹和兄弟,以及女儿和儿子是由性别("女"与"男")、代(母亲和父亲是"上一代",姐妹和兄弟是"同一代",而女儿和儿子则是"下一代")来区分的。

通过 CA 得到的深刻见解被认为揭示了不同社会的人们是如何建构各自独特的有关世界的知识和诠释他们的生活经验的。因此,CA 被认为是寻访文化和语言相对论性质的主要工具之一。

CA 的支持者旨在令它为我们提供一种有洞察力的、精妙的和严格的确定词汇含义的方法。所谓有洞察力是指它揭示了在意料之外的构成有意义的关系的基础,所谓精妙的是指它展示了其他方法所没有的拙朴和规范,而严格则是指它给我们提供了精确的定义。大多数词典的定义都由这样的特性组成(例如,跋涉"以费力和疲倦的方式行走")。

CA 的支持者也希望它能提供一组普遍的、能允许我们进行任何语言的任何含义的分析特性,这样便可克服在两种语言之间进行翻译时存在的不精确性。译者可以通过加一些短语对丢失的有意义的特性作解释,对不同语言之间存在的语义差异作出补偿。例如,在另一种语言有几个不同的词来表达用特定的方式携带(carry)什么东西时,我们可以加一些携带一词的短语,如"在手中""在臀部""在胳膊上"和"在肩上"等短语来表达这些差异。

从 1950 年代中到 1970 年代这一时期,CA 在认知人类学【Cognitive Anthropology】这一学科的发展中起到了重要的作用(D' Andrade, 1995)。然而它却因为某些原因受到了学术界的抵制。第一,它允许进行备择分析,但却无法对每一种分析究竟在多大程度上反映了文化内部的人的知识这一问题作出评价。第二,它不能对范围广泛的领域的意义作出解释。例如那些反映部分整体关系的词,如作为身体一部分的"头",也许只有在联系它在整体中的相对位置,而非列出几种特性时才能被理解(Fillmore, 1975)。第三,那些没有清楚的边界定义的类别则无法用 CA 进行分析。诸如这样的例子有生命的阶段(婴儿、儿童、青少年、成年和老年)和大小(沙子、碎石、鹅卵石、石

头、礁石和岩石)。第四,许多词的意义只能依据较为广阔的文化背景被理解。例如民族提琴(fiddle),它可以与小提琴(violin)有所区别,是因为它可以与吉他、夏威夷四弦琴和五弦琴组成乐队,为身着背带工装和格子呢衬衫的,跳着活泼的土风舞的男人伴奏。而小提琴则可和其他的提琴,一把中提琴和一把大提琴一起,作为弦乐四重奏的一部分组成一个四重奏乐队,在室内乐会上,由身穿礼服的演奏者演奏。

CA 的不朽价值在于它强调了这样一个事实——人们区别事物的根据是它们究竟相同还是不同。

——Kenneth A. McElhanon
Thomas N. Headland
(沈崇麟译校)

*也可参见认知人类学【Cognitive Anthropology】、主位/客位辨析【Emic/Etic Distinction】。

## 参考文献

D'Andrade, R. (1995). *The development of cognitive anthropology*. Cambridge, UK: Cambridge University Press.

Fillmore. C. (1975). An alternative to checklist theories of meaning. In C. Cogen, H. Thompson, G. Thurgood, K. Whistler, & J. Wright (Eds.), *Proceedings of the First Annual Meeting of the Berkeley Linguistics Society* (pp. 123-131). Berkeley, CA: Berkeley Linguistics Society.

Nida, E. A. (1975). *Componential analysis of meaning: An introduction to semantic structures*. The Hague, The Netherlands: Mouton.

# 计算机模拟(Computer Simulation)

计算机模拟使用一组方程式、算法【Algorithms】或电子元件来复制一个比较复杂的系统的主要成分。它的目的是使我们能够对现实世界中操控困难、昂贵或不可能的复杂系统进行研究。计算机模拟也允许我们在数字上对数学方程进行研究,即使这些方程用代数法是无解的。

大多数计算机模拟都是在数字计算机上进行的,并用方程、if-then(如果-那么)语句和变量来设定。各种专门的计算机语言和软件包被开发出来用于模拟的设定,尽管模拟也可以用多用途的编程语言,如 FORTRAN 或 C++来写。模拟的性状将随方程和规则中使用的参数【Parameters】、初始化模拟使用的值和在特定模拟运行时随机变量的值的变化而变化。

大多数计算机模拟都属于这三种类型:系统动态模拟(systems dynamics simulations)、有限元模拟(finite element simulations)和基于代理的模拟(agent-based simulations)。

系统动态模拟为福里斯特(Forrester, 1961)所首创,它们都为一组差分方程所驱动。而这些方程则通过一组反馈回路和随机干扰与系统中的变量相联系。罗伯特等人(Roberts, Andersen, Deal, Grant, & Shaffer,1983)则对之作了扩展。例如"全球动态"模拟研究具有高度综合的变量,如总人口水平、资源消费、污染和经济生产。当生产和消费增长的时候,污染就会加剧,而这随后又会降低人口数量的水平。

与此相反,有限元模拟通过观看模拟中相邻元素之间发生的微观水平过程的效应,计算综合变量的性状。这种方法一般用来为物理系统(如气候)建模。因为在气候系统中虽然宏观水平的方程是已知的,但因为随机变量的存在,系统在代数上是非线性【Nonlinear】的。例如,气候变化模型包含调节温室气体,如大气中取暖排放的二氧化碳

的效应、相邻的海洋温度对大气热量的效应、海洋上方大气云的形成对海洋温度的效应及其他效应的方程。大多数有限元模拟都能在含有数以百计或千计的单个处理器的巨大并联计算机上进行，而这些“超级计算机”的应用几乎都是有限元模拟。

最后，基于代理模拟则通过观看性状由规则设定的个体代理的综合效应来为社会系统建模(Axelrod, 1997)。例如，城市交通流量模拟，“代理”就是单个的汽车，这些汽车可以改变自己的车速、行车路线和随机行为(如根据其他代理的行为判断遇到事故的可能性而改变的行为)。有关这一方法的更为详细的介绍请参见模拟【Simulation】这一词条。

虽然大多数社会系统的模拟是用数字计算机做的，但仿真计算机模拟有时也会用于社会科学中，特别是在神经系统研究方面和诸如模式识别这样的研究中的建模。仿真计算机是用诸如电阻、电容和晶体管这些元件构建的。与数字模拟一样，仿真计算机是比较复杂的系统的简化，但它的特性操控起来较之用全系统的模拟更为容易和可靠。

——Philip A. Schrodt

(沈崇麟译校)

### 参考文献


Axelrod, R. (1997). *The complexity of cooperation: Agent-based models of competition and collaboration.* Princeton, NJ: Princeton University Press.

Forrester, J. W. (1961). *Industrial dynamics.* Cambridge: MTT Press.

Roberts, N., Andersen, D. F., Deal, R. M., Grant, M. S., & Shaffer, W. A. (1983). *Introduction to computer simulation: A system dynamics modeling approach.* Reading, MA: Addison-Wesley.


## 计算机辅助数据收集(Computer-Assisted Data Collection)

这一术语可用于几种不同的情境：(a)面对面和电话访谈，调查员根据访谈安排【Interview Schedule】的预定，对置于计算机内的一系列问题进行访谈，并在访谈进行的同时将回答人的回答输入计算机。(b)自填式问卷【Self-Administered Questionnaires】，回答人自行阅读屏幕上显示的问题，并于此时将自己的答案输入计算机。(c)网络调查【Internet Surveys】。(d)在线焦点小组【Focus Group】和个人访谈。

——Alan Bryman

(沈崇麟译校)

## 计算机辅助个人访谈(Computer-Assisted Personal Interviewing)

计算机辅助个人访谈(CAPI)和它的“同宗兄弟”，计算机辅助电话访谈(CATI)主导着调查数据的收集(参见计算机辅助数据收集【Computer-Assisted Data Collection】)。目前可用于这方面的计算机软件非常多，且今后还有更多的会陆续上线。本词条不准备评价这方面的各种备择方法，而是要讨论CAPI系统需要解决的主要问题，进而给读者提供解决的大致方案。计算机辅助的优点实在太多，如高效的问题筛选、值域检查和跳转式检查(skip pattern checks)等，不胜枚举，以至于不用计算机辅助方法的调查已

十分罕见。除非在遇有问卷十分简单和只涉及少数几个回答人的调查,或可以使用报酬很低的人工处理数据和无法提供硬件时,我们才使用非计算机辅助调查方法。

设计组织一个重要的抽样调查【Survey】需要协调许多方面的工作和详细说明有关各项数据收集的事宜。设计一个CAPI软件系统来处理大型、复杂的调查是很困难的。除非CAPI系统和调查过程其他方面整合得很好,否则就不只是计算机化的数据收集的效益会被抵消,而且低劣的系统整合还会造成各种意想不到的困难。CAPI首先承诺通过内置的自动检验,防止或纠正误差和不一致性。此外,它还承诺加快从实地调查结束到数据发布的数据传递时间。CAPI的确实现了自己的承诺。但在另一方面,为调查付钱的那些组织希望CAPI降低调查的费用。一般来讲,这一希望似乎落空。我们认为,费用方面的问题反映我们未能成功地整合CAPI软件和调查过程的其他方面的工作。

组织实施一个重要调查有六个彼此有着紧密联系的工作要做。第一,我们必须设计好整个访谈,以满足项目的需求。第二,必须准备好把问题和正当的答案放在屏幕的适当地方的软件。第三,准备好任何用软件进行访谈所需的输入数据。输入数据多少不等——也许可能只是一些初步的基本数据(因为调查只有少量,甚至没有涉及将要被调查的回答人的数据);但也许数量巨大而复杂,含有来自前几轮调查的成千上万个数据(它们可作为本次调查的参考)。例如,在收集事件史【Event History】数据时,家庭成员的姓名、人口学特征、雇主姓名、日期和转换的类型等可以是构成一个纵向调查的输入数据文件的主要内容。输入数据也可以包括供计算机在向回答人"宣读"问题的文本时使用的声音文件,这种方法常用于有关性行为和性器具这样一些直接由调查员问回答人,答案可能产生偏倚的敏感问题。第四,培训调查员使用软件,并派他们出去收集或"捕捉"数据。第五,将数据从调查员的袖珍PC上转移到中心数据存储库。第六,整理中心数据库中的数据准备好分析的数据文件。

虽然本词条的剩余部分将集中在软件问题,但是任何用于CAPI的软件都需要在硬件上运行。目前,用于这方面的硬件主要有三类:笔记本电脑(laptops)、手写电脑系统(pen-based systems)和小型掌上个人数码助手(PDA)。对大型的复杂调查而言,首选的硬件是笔记本电脑,因为它是一种比较成熟的产品,有其他同类产品不可比拟的多样性、稳健性和稳定性。不过基于视窗/英特尔(Windows/Intel)技术的手写电脑系统也占有一定的市场份额。早期的手写电脑系统性能似乎不够好。新一代的视窗/英特尔手写电脑性能大大优于早几代的手写电脑,但是对于众所周知的节约成本的行业,价格、可靠性和物理稳健性则是接受将新一代的手写PC机用于调查的关键所在。对于站着进行的简短调查,新型的有着更强的处理能力和较大的储存空间的掌上个人数码助手则是一位较强的竞争者,但是它还不够成熟的操作系统和数量十分有限的应用软件则阻碍了它的广泛应用。然而,在那些体积小、质量轻和低成本为先决条件的应用中,PDA则已经成为首选的一种硬件。

在编写用于数据收集的软件时,系统设计者往往专注于数据捕捉的步骤。他们经常对系统设计将可能对调查员的培训发生什么影响,以及数据捕捉软件如何与其他五项工作整合注意不够。随着CAPI的出现,调查工作的整个后勤部门的运作必须重新进行设计。如果没有将CAPI数据收集的主要工作整合起来的、统领一切的概念框架,那么重新设计调查过程的成本常常会意想不到地高。

CAPI/CATI 法有两个主要的分支。使用比较普遍的那种方法是用自定义的编程语言来提出问题、捕捉答案和储存数据。这种方法在调查过程中需要得到编程员的大量支持，而在每开始一个新的调查时，调查过程涉及的所有这些工作的相互关系必须重新设定。我们把这种方法叫作编程法（programming approach）。其他方法的 CAPI/CATI 软件使用关系数据库法作为一种整体化的机制。在使用这种方法时，工具被表示为一系列数据记录，每种数据记录都有很多属性，包括问题文本、文档跳答模式注释和给调查员的指令等几乎所有的东西。我们把这种方法叫作数据库法（database approach）（读者可参阅有关数据管理【Data Management】的词条，该词条对数据库作了清楚的解释，它与本部分的内容关系也很密切）。在这里我们要说明的问题是数据库给我们提供的整合 CAPI 调查的所有步骤的提纲挈领式的思路，这使它比备择的编程法在操作上更为经济。

在设计阶段，调查专家将数据输入屏幕的各个字段。这些字段位于代表问卷的关系数据库表的行。输入的数据包括问题的文本、给正当回应设立的校订限制，正当答案清单和转移指令。这种“编写”CAPI 问卷的方法需要一些编程技巧，因而一个初级的调查专家一般可输入那些最为复杂的段落之外的所有段落。数据库中一组预设的“查询问题”会在数据输入时对输入的数据进行诊断性检查，至于更复杂的数据检查和诊断则以批处理模式进行。工具中的每个问题则由数据库表连成的行表示。调查软件执行这些连接、显示问题、储存答案和按分支指令进行分支。每次调查宣读和实施问卷数据的程序不是重写的，而是稳定的和可以用于不同调查的（同样的记录也可用于 PDA、互联网调查，或其他能使用各种数据收集平台的本地服务器上）。这将大大简化问卷作者和调查员的培训工作，因为所有与系统的互动都会再次使用几个为数不多的标准屏显。

调查数据被转移到中心办公室，并下载到已经含有定义问卷的表格的主数据库。这时问卷数据变成了“元数据（metadata）”或描述调查数据的数据，大大便于文档报告的产生，这些报告将对关系数据库进行一系列查询。预设的查询既可产生可打印的也可产生超文本的问卷或编码手册。其他与每个问题有关的信息也可储存在数据库中，并可用于增加调查的文档。

我们可以从关系数据库中提取输入的数据用于纵向调查的下一轮调查。关系数据库技术有利于数据共享，既可用于商业用途也可用于学术用途。研究者可通过适当的途径，到数据库寻找需要的文档和变量，并加以标志，要求远程的服务器通过电子邮件将它们发给自己。

这种路数是整体性的，因而要使用商业数据库软件来整合一个 CAPI 调查的所有阶段。一个设计精良的系统能处理各种调查，允许中心办公室和调查人员再次使用标准工具。这种路数给我们提供了充分的机会来控制费用和保持运作的协调和高效（参见数据管理【Data Management】）。

——Randall J. Olsen

（沈崇麟译校）

## 参考文献


Costigan, P., & Thomson, K. (1992). Issues in the design of CAPI questionnaires for complex surveys. In A. Westlake, R. Banks, C. Payne, & T. Orchard (Eds.), *Survey and statistical computing* (pp.47-156). London: North Holland.

Couper, M. P., Baker, R. P., Bethlehem, J., Clark, C. Z. F., Martin, J., Nichols, W. L., et al. (Eds.). (1998). *Computer assisted survey information collection*. New York: John Wiley.

Forster, E., & McCleery, A. (1999). Computer assisted personal interviewing: A method of capturing sensitive information. *IASSIST Quarterly*, *23*(2), 26-38.

Saris, W. E. (1991). *Computer-assisted interviewing.* Newbury Park, CA: Sage.

## 概念(Concept)

参见概念化，操作化和测量(Conceptualization, Operationalization, and Measurement)。

## 概念化，操作化和测量(Conceptualization, Operationalization, and Measurement)

研究始于一个"问题"或题目。对一个问题的思考会导致确定一些能够捕捉这一研究现象的概念。概念，或者称构念【Constructs】，是指那些表征了某一现象的想法。概念化是赋予这些概念理论化意涵的过程。这个过程的典型做法是用理论化的术语抽象地描述概念。

描述社会现象，验证研究假设需要对概念进行操作化。操作化将研究者从抽象层次转移到经验层次，经验层次的关注点不是概念，而是变量。它涉及那些能够测量概念的操作或程序。测量是将数字符号(或某些其他的标签)附于变量的层次或特征的过程。于是实际研究就会因此成为在经验上的研究变量，并对概念作出(描述性的、关联性的或因果性的)陈述。

### 概念化

尽管研究可能只是始于几个有时是界定非常不严谨的概念，但研究的早期阶段却通常都是对概念进行界定和明确概念之间的关联。在探索性研究中，其研究目的之一常常就是更好地界定某个概念，或确定其他的重要概念和可能的关系。

由于社会科学中的许多概念都是用日常生活对话中常用的词语来命名的，对这些概念的含义进行界定就显得非常重要。例如，规范、不平等、贫困、公正、合法性等概念，都是我们日常生活中经常使用的词语，对于不同人而言，其含义各不相同。我们通常都把研究者所提供的界定叫作表征性界定(例如用其他的词汇来界定某个概念)。没有人声称这些界定就代表了概念的"真实"含义。这些界定的目的是当使用该概念时，我们能够更好地告知他人这个概念的意思是什么。社会科学的一个重要目标是将核心概念的界定进行标准化。有种说法认为社会科学界最为成熟的研究项目即那些产生了最多知识的研究——使那些所有人在核心概念上有一致的看法，并用相同方式进行界定的研究，这种说法值得论证。

不同概念的抽象程度是不同的。例如，人力资本是一个非常抽象的概念，而教育(通常作为人力资本的一个指标)这个概念就没有那么抽象。但是教育在质量和数量方面可以有不同变化，因而就比正式上学的年限这个概念更加抽象。大家会将社会科学中那些更为抽象的理论看作对推进知识的发展更为有用。但是，概念越抽象，在对于概念进行合适测量方法方面达成一致就越困难。

## 操作化

对概念的操作化是指由抽象层次转移到经验层次。过去社会科学研究者并不经常使用这一术语,主要因为在某些语境下,这个词的使用具有负面含义。这方面研究中的一个例子是关于人类智力的研究。由于对智力这个概念的含义难以达成一致,一些研究者干脆宣称,智力就是智力测试的结果。这样,智力这个概念就是以测量它的操作来定义的。这种所谓的"天然经验主义"受到了广泛批评,因此,很少有研究者现在通过概念如何被操作化来对概念进行定义。研究者一般会使用前文中描述过的表征性定义,而把对概念的测量看成一个独立的不同的活动。研究者虽然意识到测量并不能完美地代表概念(下文中会对此作出描述),但是研究者的目标仍然是得到有效和可信地捕捉到概念的测量。

## 测量

测量是指将数字符号(或其他标记)分配给变量的水平或特征的过程。在从概念的抽象层次转移到变量的经验层次时,研究者必须将概念设想为具有可以在经验上观察或评估的特征。许多问题必须加以注意。其中第一项是测量层次【Level of Measurement】。

在最基本的层面上,研究者必须决定某个概念的基本特征是否允许对个案进行排序(定序水平),或只允许对个案进行分类(定类水平)。另一个区别是关于概念的功能是离散的还是连续的,以及是否有精确的刻度。依托这些区别,大多数研究人员所面临的变量是无法排序并离散的(如婚姻状况、宗教偏好),有序并离散的(如社会阶层),或有序并连续的(如收入)。过去几十年统计的发展已经大大增加了有序与离散变量相关的分析工具,但社会科学的大多数研究仍然坚持分类变量与连续变量的二分法,只有更多"自由主义"倾向的研究人员会假定,定序数据是可以被看作连续变量来进行分析的。然而,随着统计技术的进步,研究者不再需要作出这些决定;在复杂的多变量研究中,定序变量还可以作为定距(此处原文为定序似为定距之误——译者注)变量来处理(Long, 1997)。

因而,测量的第一步是决定概念中固有的测量水平。一个通用的规则是,测量应始终代表一个概念的可能的最高水平。这样做的原因是显而易见的。一个连续变量在以定类的方式测量后,是不能轻易地转换成一个连续的测量尺度的。而一个连续的测量尺度总是可以转化为一个定序或定类的测量。此外,定序和连续的测量尺度允许使用更强的统计检验,也就是在应当拒绝虚无假设的时候,它会有更高的被拒绝的可能性。

任何测量通常都会涉及某种类型的测量误差【Error】。一种类型的错误是无效测量,它指的是测量错误地捕捉概念的程度(DeVellis,1991)。无效被称为系统误差或偏倚。我们通常可以这样理解效度【Validity】,即将概念作为靶心,而测量是用于击中靶心的箭头。测量命中靶心(概念)中心区域的程度就是测量的效度。研究人员希望获得效度高的测量。原因很简单,我们以一个社会流动的研究为例。如果研究人员的目标是确定某个社会中社会流动的程度,而测量从始至终低估了流动性,那么这样测量得出的结果是有偏倚和误导性的(即它们是无效的)。

测量误差的另一主要类型是不可靠。如果无效指的是测量的精确性,不可靠则是指测量在重复使用的情况下所产生的不一致性。平均而言,测量应该给出一个变量上的个案的正确分数,但是如果对于同一个个

案,在测量被再次使用时产生了不同的分数,那么该测量会被认为是不可靠的。幸运的是,不可靠可以通过统计方法被修正;而没有效度则不那么容易被修正,并且也不像信度【Reliability】那样容易评估和认定。

比起信度,我们认为效度是一个测量更为重要的特征。一个没有捕捉到概念的可靠测量几乎毫无价值可言,而基于它的使用得到的结果将会是一种误导。严谨的概念化过程对提高测量的效度至关重要。如果概念的界定就是模糊不清的,那么测量的效度很可能会受到影响,因为研究者确定不了究竟应该测量哪些东西。评估测量效度有以下几种方法:

表面效度【Face Validity】,即测量是否在“表面上”捕捉到了概念。表面效度是不规范或无法量化的,尽管如此,它却总是研究中要作出的有价值的首要的评估。例如,衡量一个家庭的家庭收入,如果只包含男性户主的收入可能会提供一致的结果(是可靠的),但其有效性则必定受到质疑,尤其是在一个有这么多双薪家庭的社会中。

内容效度(content validity)类似表面效度,但它的使用是有更严格的标准的。对于一个有内容效度的测量,它的界定是必须能像它所定义的那样,捕捉到概念的所有维度或特征。例如,一个一般工作满意度测量应包括薪酬的满意度、工作保障的满意度、晋升机会的满意度等。但是,与表面效度类似,内容效度也是几乎无法量化的。

效标效度【Criterion-Related Validity】是某种测量与其他被认为是一个概念的精确指标的量度之间的相关程度。效标效度有两种形式:预测效度,采用一种未来的标准。例如,投票偏好(在选举前测量)与实际投票行为有相关性。高相关性表明,投票偏好是投票行为的一个有效测量。同时效度是通过获得某个测量,与另一个已被证明是概念的有效指标的测量之间的相关性来评估的。例如,工作场所社会学的研究者,希望通过询问员工参与决策的程度来测量员工的参与问题。有一个更费时的策略——视频拍摄工作组的会议,并记录每个员工的贡献——该测量已被视为对参与的有效测量。这两者之间的相关将表明这种感性测量的同时效度。

一种测量的构念效度【Construct Validity】是指两个有效性评估策略中的某一个。首先,它可以指当一个变量用该测量进行评估时,其体现的性状,是否如其应该体现的那样。例如,如果理论(和/或过去的研究)认为它应该与另一个变量 $Y$ 有正相关关系,那么这种关系应该在使用该测量时被发现。其次,概念的多个指标都与某个基本的概念建构,而不是其他概念建构之间相关的程度。通常情况下,这种区别被称为聚合效度和区分效度【Discriminant Validity】。因子分析【Factor Analysis】可以用来评估这一点。例如,如果一个研究者有五个文化资本的指标和四个社会资本的指标,因子分析应该产生两个低相关的因子,每个因子分别属于两个指标组。

需要强调的是,这第二个策略运用因子分析来实现验证的目的,而不是用来进行数据梳理或概念发掘,这种使用法多发生在探索性研究中。如果研究人员在概念化过程中一直非常严谨,那么因子分析便可用来验证他们的测量。在这种情况下,因子分析不再是用来“发现”未预料到的概念了。

综上所述,信度是指当测量被重复使用时其结果的一致性。它被看作随机测量误差,而效度被认为是测量偏误。没有信度的测量会造成各种显而易见的统计后果,其中最常见的是方差的扩大和相关系数的缩小。信度通常都用以下两种方式中的一种来评估。再测信度【Test-Retest Reliability】是用同一测量的两个时点之间的相关系数来评估的。这种类型评估的困难在于,测量的不

可靠性会与测量的变量的实际变化发生混淆。出于这个原因,第二种也是更加常用的策略,是获得概念的多个指标。在这里,可以用内部一致性测量,如 Cronbach α【Cronbach's Alpha】系数。在这两种情况下,规范的关联(相关)测量允许使用某些经验法则。例如,任何低于 0.60 信度的测量都应该被看作存在着相当大的可疑之处。

## 其他问题

测量效度不应该与内在效度【Internal Validity】和外在效度【External Validity】相混淆,后者属于研究的设计问题。如果研究人员声称研究结果支持他们的 *X* 导致 *Y* 的假设,且该结果的备择的解释已经被排除,那么我们就可以说该结果有内在效度。外在效度是指研究结果是否能够推广到研究的场景、样本或时间框架以外。内在效度和外在效度都是构建社会科学知识中非常重要的问题,但它们与测量效度却不尽相同。

另一个与测量有关的问题是,测量究竟应该使用单一指标,还是多个指标。一些概念,从定义几乎就可以看出,是能够用单一指标来测量的,例如收入和教育。而其他较抽象的概念(例如,政治保守主义和婚姻满意度)经常需要用一个以上的指标来测量。需要多个指标的原因是显而易见的。第一,当一个概念是更抽象的,找到能捕捉该概念的单一指标是比较困难的。第二,多个指标的测量通常比单一指标的测量更有信度。第三,在结构方程模型【Structural Equation Modeling, SEM】中(Bollen, 1989)使用多个指标的测量,使我们得以对无信度的测量进行修正,产生额外的信息,从而使我们得以进行有用的模型检验和估计更复杂的模型,例如交互效应模型。尽管结构方程模型能够通过分别处理每个指标来加强它的效用,但基于多指标的测量常常会被转化为指标简单加总,或加权加总的尺度或指数。

第三个与测量有关的问题是有效和可信的测量的来源。正如前文已经描述的那样,有多种策略可用来评估某个测量究竟有多好(即它是多有效和多可信)。但是,这种评估发生在测量后。如果某个测量没有效度或信度,而研究者又无法轻易获取更好的测量,那么就应该认真考虑研究的中止。出于这个原因,我们应当重视和认真考虑在研究的开始阶段识别有效和可信的测量。如果在一个专业领域,理论已经完备建立,且已经有了一个强大的科研传统,那么有效和可信的测量很可能也已经存在。例如,在组织的研究中,像形式化和行政力度这样的关键概念已经非常成熟,关于这些概念的测量也是如此。在这种情况下,仔细阅读该领域的文献,便能够用来确认需要使用的测量。对于那些进入某个专业领域的新人,希望“找到一个更好的测量”的想法是很诱人的,但除非先前使用的测量具有较低的信度和效度,研究者应该远离这种诱惑。在关键概念的测量跨研究地被标准化时,知识的积累才更有可能。当然,探索性研究不太可能使用标准化的测量。在这种情况下,概念和测量会“沿路”发展。但是,这样的研究不应该被看作是完全非结构化、完全没有评估和解释的标准。马克斯韦尔(Maxwell, 1996)令人信服地表明,探索性的研究仍可以在准科学的指导准则下进行。

最后一个与测量有关的问题涉及对现有数据的利用,现有数据是社会科学研究的主要来源。通常情况下,研究人员收集这些数据时已在考虑其他的一些目的。在这种情况下,经常需要依靠代理变量【Proxy Variables】,并且必须面对效度问题。信度往往不是什么问题,原因是代理变量似乎通常都是社会人口变量。幸运的是,现在的大型数据收集机构收集数据时,在测量问题上

都已经更多地倚重社会科学家。必须确定和使用关键概念的代理变量这个问题有望成为过去。

——Charles W. Mueller

（郑晓娟译　沈崇麟校）

### 参考文献


Bollen, K. (1989). *Structural equations with latent variables*. New York: John Wiley.

DeVellis, R. F. (1991). *Scale development: Theory and applications*. Newbury Park, CA: Sage.

Long, J. S. (1997). *Regression models for categorical and limited dependent variables*. Thousand Oaks, CA: Sage.

Maxwell, J. A. (1996). *Qualitative research design*. Thousand Oaks, CA: Sage.


## 条件似然比检验(Conditional Likelihood Ratio Test)

条件最大似然比是一种对用条件最大似然估计【Conditional Maximum Likelihood Estimation】的参数上的约束进行检验的程序。寓于条件似然比检验统计量背后的直觉知识与似然比统计量【Likelihood Ratio Statistic】类似。条件似然比统计量测量强制约束的条件似然比函数的约化。如果约束是正当的,条件对数似然比的约化便是不显著的。大样本的条件似然比统计量的分布为自由度【Degrees of Freedom】等于强制约束数的卡方分布。与似然比统计量一样,条件似然比检验也有对应的瓦尔德和得分(拉格朗日乘子)检验(Wald and score [Lagrange multiplier] tests)的条件版。更为正式的表达是,假设我们有一个用最大化下面的条件对数似然比函数,$\ln L = \sum f(y_2 \mid \mathbf{X}, \boldsymbol{\beta}, \mathbf{Z}, \gamma)$估计的参数向量 $\boldsymbol{\beta}$,式中的 $\gamma$ 是一个参数向量,其值为理论假设设定,或更为普遍的是由最大边际对数似然函数,$\ln L = \sum g(y_1 \mid \mathbf{Z}, \gamma)$ 求得的估计值 $\hat{\gamma}$。现在我们来考虑一下 $\boldsymbol{\beta}$ 的$(r\times1)$的约束向量,将它定义为 $\mathbf{c}(\boldsymbol{\beta})-\mathbf{q}=\mathbf{0}$。条件似然比检验以 $\ln \hat{L}, \boldsymbol{\beta}$ 的无约束条件对数似然函数,与 $\ln \hat{L}_R, \boldsymbol{\beta}$ 的约束对数似然比函数之差为依据。注意,二者都是在 $\gamma$ 上有条件的,而 $\gamma$ 则是作为已知处理的。然后,条件似然比统计量则是$-2(\ln \hat{L}_R-\ln \hat{L}) \sim \chi_r^2$,它对作为虚无假设的 $\mathbf{c}(\boldsymbol{\beta})-\mathbf{q}=\mathbf{0}$ 进行检验。

条件似然比检验的一种应用,是在联立概率单位模型(simultaneous probit models)情形中的外生性(exogeneity)检验。我们来看一下概率单位模型的一个特例。该模型中的一个回归量,$y_1=\mathbf{Z}\gamma+v$ 是一个连续的有着正态分布【Normal Distribution】的内在变量【Endogenous Variable】。这一方程系统的条件最大似然估计涉及:先要用普通最小二乘【Ordinary Least Squares】求 $\gamma$ 的最大似然估计值,然后再将这一估计值用于调节概率单位方程的最大似然估计。在这种情形下的调节是用 $\hat{\gamma}$ 导出约化式的残差 $\hat{v}$。这一残差包括在概率单位方程的右边,控制方程联立偏倚。条件似然比检验将这种未约束的模型的对数似然,与用排除简化式残差的强加外生性约束的约束模型的对数似然进行比较。在这一特例中,条件似然比统计量有自由度为 1 的卡方分布。注意,在外生性虚无假设下,这一似然比检验渐近地等于似然比检验。

——Harvey D. Palmer

（沈崇麟译校）

## 参考文献

Alvarez, R. M., & Glasgow, G. (2000). Two-stage estimation of nonrecursive choice models. *Political Analysis*, *8*, 147-165.

Rivers, D., & Vuong, Q. H. (1988). Limited information estimators and exogeneity tests for simultaneous probit models. *Journal of Econometrics*, *39*, 347-366.


# 条件 Logit 模型(Conditional Logit Model)

条件 logit 是一种研究有着三类或更多类的无序分类【Categorical】(定类【Nominal】)因变量问题的统计模型。条件 Logit 估计每个观察将在因变量的每一类中的概率。条件 Logit 使用选择类别的特性作为自变量,因此不允许自变量的效应跨选择类别变化。这一点不同于类似的用个别观察的特性作为自变量,且允许自变量的效应跨因变量类别变化的多项 Logit【Multinomial Logit】模型。模型同时包括备择选择的特性和个别观察的特性二者是可能的,而这样的模型通常也叫作条件 Logits。注意,有些研究者将多项 Logit 和条件 Logit 并称为"多项 Logit"。

有几种不同的导出条件 Logit 模型的方法,在社会科学中使用最为普遍的方法是,将条件 Logit 模型作为一个离散选择模型导出。每个观察都是行动者(actor)从选择的一组 $J$ 个备择选择中选出的一个备择选择。在这里,$J$ 是因变量中的类别数。一个行动者 $i$ 从这一选择组选出的备择选择 $k$ 将要得到的效用(utility)由下式给出:

$$U_{ik} = X_{ik}\beta + \varepsilon_{ik}$$

令 $y$ 为有 $J$ 个无序分类的因变量。那么,一个行动者 $i$ 在备择选择 $k$ 上将要选出备择选择 $j$ 的概率便是

$$\begin{aligned}\Pr(y_i = k) &= \Pr(U_{ik} > U_{ij}) \\ &= \Pr(X_{ik}\beta + \varepsilon_{ik} > X_{ij}\beta + \varepsilon_{ij}) \\ &= \Pr(\varepsilon_{ik} - \varepsilon_{ij} > X_{ij}\beta - X_{ik}\beta)\end{aligned}$$

麦克法登(McFadden,1973)证明,如果所有的 $\varepsilon$ 都是独立的,且服从跨所有观察和备择选择的 Gamble 分布(Gumbel distributions),那么这一离散选择导子(derivation),将导出一个条件 Logit。多项 Logit 估计 $\Pr(y_i = k \mid x_{ij})$,式中的 $k$ 是因变量中的 $J$ 个分类中的一类。这个条件 Logit 模型写作:

$$\Pr(y_i = k \mid x_{ik}) = \frac{e^{x_{ik}\beta}}{\sum_{j=1}^{J} e^{x_{ij}\beta}}$$

条件 Logit 是用最大似然法【Maximum Likelihood】估计的。与多项 Logit 不同,这些变量的系数都不必确定约束。然而,那些跨备择选择不变的变量的系数则必须以某种方法加以约束,以对模型进行识别。最为普遍的约束法是将跨个体不变的各个变量的 $\beta$ 正态化,以使一个备择选择等于零。

## 不相干备择选择的独立性

条件 Logit 和多项 Logit 有一个叫作不相干备择选择的独立性的性质(IIA)。对于每个观察者而言,在任何两类中的概率比都不受来自选择组其他类别的添加或移除的影响。这是因为任何两个备择选择的选择概率的比只取决于适合这两个备择选择的变量和系数。为了了解这一点也适用于条件 Logit,我们来看一下下面的式子:

$$\frac{\Pr(y_i = m \mid x_{im})}{\Pr(y_i = k \mid x_{ik})} = \frac{e^{x_{im}\beta} \Big/ \sum_{j=1}^{J} e^{x_{ij}\beta}}{e^{x_{ik}\beta} \Big/ \sum_{j=1}^{J} e^{x_{ij}\beta}} = \frac{e^{x_{im}\beta}}{e^{x_{ik}\beta}}$$

注意,IIA 是个别观察的选择概率的一种性质,而非用于模型总概率的一种性质。在许多设定中,这一性质常常是不可取的。说明这一性质最为常用的一个例子是红色公共汽车/蓝色公共汽车悖论。假设某人去上班的时候,既可选择开车,也可选择一辆刚好油漆为红色的公共汽车。令这个人选择每一备择选择的概率为 1/2。现在第二辆恰巧油漆为蓝色的公共汽车加入了选择组,蓝车除了颜色以外,其他每一方面都与红车相同。直觉告诉我们,开车和每一种公共汽车的选择的概率应该是 1/2 和 1/4。然而 IIA 则需要每种交通方式的概率都为 1/3,以保持在开车和乘公共汽车之间选择的原概率。注意 IIA 的这一性质适用于任何假定误差项的分布是独立和同一的离散选择模型。而希望放松 IIA 这一性质的愿望便成了一个导致在社会科学中的多项选择问题研究中,诸如多项式概率(MNP)这样一些模型使用增长原因之一。

## 例子

作为一个例子,我们来考虑一个在一次有三个候选人(A,B 和 C)的选举中,年龄、性别、收入和意识形态距离对候选人投票选择的效应的估计问题。年龄、性别和收入是跨备择选择不变的变量,因此它们将作为多项 logit 中的变量处理。这就是说,这些变量的系数将跨候选人变化,因而为了确定模型,它们必须以某些方法加以约束。在这个例子中,将这些变量的系数中的候选人 C 的系数约束为零。个人和候选人之间的意识形态距离是由个人自我确定的在 7 级意识形态量表的位置和由他(或她)确定的候选人在同样的量表上的位置的差别来量度的。因为这一变量是跨个人和备择选择变化的,所以意识形态距离对候选人选择的影响只有单独一个系数,因此这个系数不需要用任何方法约束。

这一选择问题的条件 Logit 估计结果在表 1 中列出。

**表 1　多项 Logit,多候选人**
**(候选人 C 的系数正态化为零)**

| 自变量 | 系数值 | 标准误差 |
|---|---|---|
| 意识形态距离 | −0.12* | 0.04 |
| | *A/C* 系数 | |
| 收　入 | −0.51* | 0.20 |
| 年　龄 | −0.01 | 0.01 |
| 女　性 | 0.08 | 0.15 |
| 常　数 | 0.43 | 0.67 |
| | *B/C* 系数 | |
| 收　入 | 0.24* | 0.09 |
| 年　龄 | −0.06* | 0.01 |
| 女　性 | 0.27 | 0.23 |
| 常　数 | 0.43 | 0.67 |
| 观察数 | 1 000 | |

注:* 在 95%水平统计上显著。

注意,年龄、性别和收入的系数是跨备择选择变化的,因而一个备择选择的这些系数已被正态化为零。这个条件 Logit 告诉我们,随着收入的增长,个人投票给候选人 A 的可能比投票给候选人 C 的小,但是投票给候选人 B 的可能则比投票给候选人 C 的大。随着年龄的增长,个人投票给候选人 B 的可能比投票给 C 的小。意识形态的系数是不跨候选人变化的,它告诉我们,随着个人和一个候选人的意识形态距离的加大,个人将较少有可能投票给那个候选人。

——Garrett Glasgow

（沈崇麟译校）

## 参考文献

Long, J. S. (1997). *Regression models for categorical and limited dependent variables.* Thousand Oaks, CA: Sage.

McFadden, D. (1973). Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Ed.), *Frontiers of econometrics* (pp.105-142). New York: Academic Press.

Train, K. (1986). *Qualitative choice analysis: Theory, econometrics, and an application to automobile demand.* Cambridge: MIT Press.

# 条件最大似然估计(Conditional Maximum Likelihood Estimation)

全信息最大似然(FIML)估计的一种备择方法是条件最大似然估计(CMLE),它通过将某些参数作为已知处理,将最大化问题简单化。CMLE 涉及对数似然函数的最大化。将参数作为已知处理的方法主要有两种:一种是通过理论假定来设定;另一种使用更为普遍的方法是用估计值来取代。CMLE 的应用涵盖范围很广,从自相关【Autocorrelation】的迭代的 Cochrane-Orcutt 法到递归联立方程【Recursive Simultaneous Equations】的两步估计法。CMLE 法产生的一致性估计值,其效率普遍低于有限信息最大似然(LIML)估计,即使如此,在某些条件下,这两种方法是渐近【Asymptotically】等价的。在全对数似然函数难以导出或最大化时,CMLE 被用来替代 FIML。因此,随着计算机计算方法的优化和速度的加快,CMLE 应用的频数已经减少。然而,作为产生最大似然估计值的迭代方法,如用于贝叶斯分析期望最大化(EM)算法的中间步骤,CMLE 还是很重要的。

CMLE 仍常用于作为一种估计递归方程系统参数,如涉及作为连续内生【Endogenous】回归量的函数的有限个因变量的参数估计的工具(例如联立的 probit 和 tobit 模型)。CMLE 并不最大化系统地联合对数似然函数,而是将系统顶部的方程的条件对数似然函数,通过用最大似然估计值替代系统下部方程的参数的方法最大化。在这种两步法中,第一步涉及最大化内生回归量的边际对数似然函数问题,而第二步则用来自第一步的参数估计值,对系统顶部的方程的参数最大似然估计【Maximum Likelihood Estimation】进行调整。

让我们更为正式地来看一下两个随机变量【Random Variables】$y_1$ 和 $y_2$ 的联合分布 $h(y_1, y_2 | \mathbf{X}, \beta, \mathbf{Z}, \gamma, \lambda)$。假设这一联合分布能分解为 $y_1$ 的边际分布,$g(y_1 | \mathbf{Z}, \gamma)$ 和 $y_2$ 的条件分布 $f(y_2 | \mathbf{X}, \beta, E[y_1 | \mathbf{Z}, \gamma], \lambda)$ 的因子,式中的 $\beta$ 是 $\mathbf{X}$ 的内生回归量的参数向量。而 $\lambda$ 则是定义 $y_2$ 和 $E[y_1 | \mathbf{Z}, \gamma]$ 之间的关系的参数。$\beta$ 和 $\lambda$ 的 CMLE 是一个分两步进行的过程。第一步,将边际对数似然函数,$\ln L = \sum g(y_1 | \mathbf{Z}, \gamma)$ 最大化。求出估计值 $\hat{\gamma}$。第二步,用 $\hat{\gamma}$ 取代 $\gamma$,将条件对数似然函数,$\ln L = \sum f(y_2 | \mathbf{X}, \beta, E[y_1 | \mathbf{Z}, \gamma], \lambda)$ 最大化。

在联立的 probit 模型的一种特定的情形中,$g(\cdot)$ 将有标准密度,而 $f(\cdot)$ 则有 probit 密度。第一步将用普通最小二乘【Ordinary Least Squares】求 $\gamma$ 的最大似然估计值。第二步将用这一估计值调节 probit 方程的最大似然估计。这个调节将用 $\hat{\gamma}$ 导

出包含在Probit方程的右边,用于控制同步偏倚的简化残差。这种两步条件最大似然(2SCML)估计量的一个吸引人的特点是,估计的结果可用于外生性检验(即 $\lambda=0$)。这种外生性检验的一种版本是条件似然比检验【Conditional Likelihood Ratio Test】。由蒙特卡洛法得到的证据表明,这种2SCML估计量优于两种备择方法:工具变量【Instrumental Variables】probit和两步probit最小二乘(Alvarez & Glasgow, 2000; Rivers & Vuong, 1988)。此外,在外生性下,这种2SCML估计量渐近等价于LIML估计量。

——Harvey D. Palmer

(沈崇麟译校)

## 参考文献


Alvarez, R. M., & Glasgow, G. (2000). Two-stage estimation of nonrecursive choice models. *Political Analysis*, *8*, 147-165.

Blundell, R. W., & Smith, R. J. (1989). Estimation in a class of simultaneous equation limited dependent variable models. *Review of Economic Studies*, *56*, 37-58.

Greene, W. H. (2000). *Econometric analysis* (4th ed.). Upper Saddle River, NJ: Prentice Hall.

Rivers, D., & Vuong, Q. H. (1988). Limited information estimators and exogeneity tests for simultaneous probit models. *Journal of Econometrics*, *39*, 347-366.


# 置信区间(Confidence Interval)

一个 $100(1-\alpha)\%$ 的置信区间是一个样本量为 $N$ 的随机抽样、围绕总体【Population】参数 $\theta$ 的区间估计值,预期它有 $100(1-\alpha)\%$ 次包括了 $\theta$ 的真值。置信区间是对用参数估计【Parameter Estimation】的总体参数的自然修饰,因为它表明了用样本统计值估计的总体参数【Population Parameter】的精度。

$100(1-\alpha)\%$ 的置信水平是预先选择的,一个双侧【Two-Sided】的区间使用下限 $L$ 和上限 $U$,每个都有 $100(1-\alpha/2)\%$ 次包含了 $\theta$ 的真值,因此二者一起共有 $100(1-\alpha)\%$ 次包含了 $\theta$ 的真值。这一区间经常写作 $[L,U]$,但有时我们也把置信水平和区间合起来写作 $\Pr(L<\theta<U)=1-\alpha$。在有些应用中,使用的是单侧的区间。各种各样的总体参数,如均值【Mean】、方差【Variance】、比例和 $R$ 方【$R$-Squared】,都可以计算置信区间。

置信区间与显著性检验【Significance Test】密切相关,因为 $100(1-\alpha)\%$ 的置信区间包含了所有显著标准为 $\alpha$,显著性检验不能拒绝的假设的总体参数值。在这一方面,它比显著性检验提供的信息更多。随着样本量变大,置信区间逐渐变窄和/或置信水平逐渐变低。

限 $L$ 和 $U$ 来自样本统计值(常常是 $\theta$ 的样本估计值)和一个抽样分布。该分布设定了得到样本统计量可取的每一个值的概率。因此,$L$ 和 $U$ 也是样本统计量,且随样本的变化而变化。例如,假设我们对一个来自大总体,由15个成人组成的均值为104,标准差(SD)为18的随机样本进行标准的IQ检验。我们将为均值 $\mu$ 构建一个双侧的95%的置信区间。限 $L$ 和 $U$ 必须有一种性质,如果显著标准为 $\alpha$、样本量为15、均值为104和标准差为18,那么我们可以拒绝的假设是 $\mu>104+U$ 或 $\mu<104-L$,而不是 $L<\mu<U$。图1显示了这样一些限。

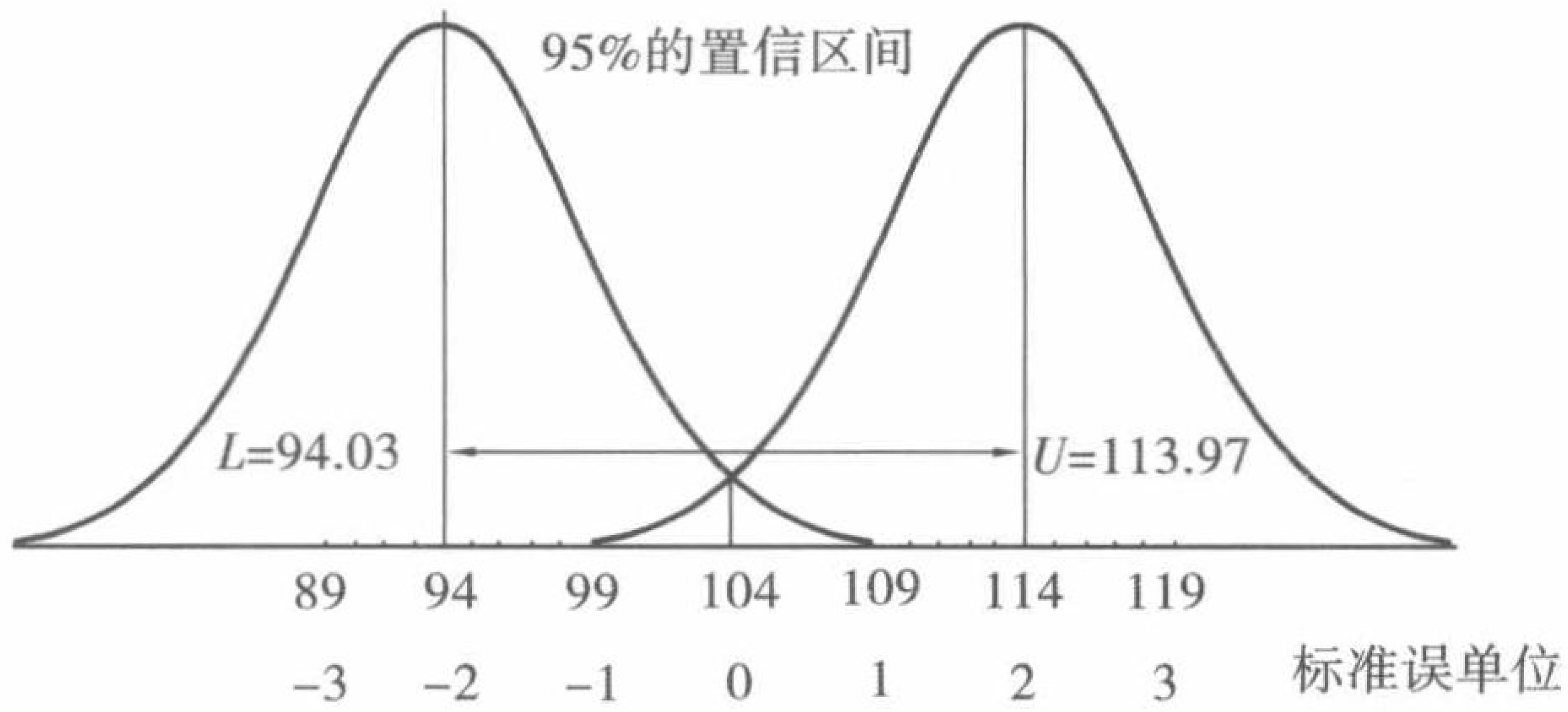



图 1 95%的置信区间的下限和上限

$t$ 比率【$t$ Ratio】的抽样分布由

$$t=\frac{\bar{X}-\mu}{s_{err}}$$

定义，式中的 $s_{err}$ 表示均值的标准误。它服从 $df=N-1=14$ 的 $t$ 分布。当 $df=14$ 时，均值之上的 $t_{\alpha/2}=2.145$ 个标准误切去了 $t$ 分布尾部的 $\alpha/2=0.025$。同样，均值之下的 $-t_{\alpha/2}=-2.145$ 个标准误则切去了 $t$ 分布下尾的 $\alpha/2=0.025$。样本标准误是 $s_{err}=SD/\sqrt{N}=4.648$。这样围绕 $U=104+(2.145)\times(4.648)=113.97$ 的 $t$ 分布有 0.025 的尾部在 104 之下，而围绕 $L=104-(2.145)\times(4.648)=94.03$ 的 $t$ 分布有 0.25 的尾部在 104 之上。这些限满足了上面要求的性质，所以 $\mu$ 的 95%的置信区间是[94.03, 113.97]。

——Michael Smithson

（沈崇麟译校）

## 参考文献

Altman, D. G., Machin, D., Bryant, T. N., & Gardner, M. J. (2000). *Statistics with confidence: Confidence intervals and statistical guidelines* (2nd ed.). London: British Medical Journal Books.

Smithson, M. (2003). *Confidence intervals* (Sage University Papers on Quantitative Applications in the Social Sciences, 140). Thousand Oaks, CA: Sage.

# 保密(Confidentiality)

这一伦理原则要求研究者除了采取匿名方式之外，还要求不能泄露任何与研究参与者有关的发现。

——Alan Bryman

（郑晓娟译　沈崇麟校）

*也可参见伦理原则【Ethical Principles】。

# 验证性因子分析(Confirmatory Factor Analysis)

验证性因子分析是一种用来检验变量【Variables】之间的共同性(commonality)假设【Hypotheses】的统计方法。作为一种多元【Multivariate】方法，验证性因子分析被用于同时检验共同构建一个测量模型的多个假设。所有这样的模型都由四个部分组成。最为基本的测量模型至少必须有两个变量或指标。指标分数的变异性归结于两个起源——潜变量【Latent Variables】和量度误差【Error】。潜变量是不可观察的，它们是

从指标之间的共同性推断出来的。量度误差是指标中未能归结于潜变量的变异性。最后,可归结于潜变量的指标分数的变异程度由负荷(loadings),即量度方程估计的系数表示。在量度方程中,指标将对假设对它们有影响的潜变量做回归。

## 变异性的来源

验证性因子分析根据在量度模型中设定的来源,将指标分数的变异性进行分解。图 1 直观地阐述了这种分解的根据。图中左侧的 Venn 图(Venn diagram)显示了能为一个量度模型设定的三个指标 $V_1$, $V_2$ 和 $V_3$ 的六种可能的来源。在这种非正规的描述中,指标用圆圈表示,每个圆圈都由三个部分组成。标以阴影的区域 $F_1$ 是每个指标与其他两个指标共有的变异性。交叉线区域是只和其余指标中的一个共有的变异性,而没有阴影和交叉线的区域则是不和其他两个指标中的任何一个共有的变异性。

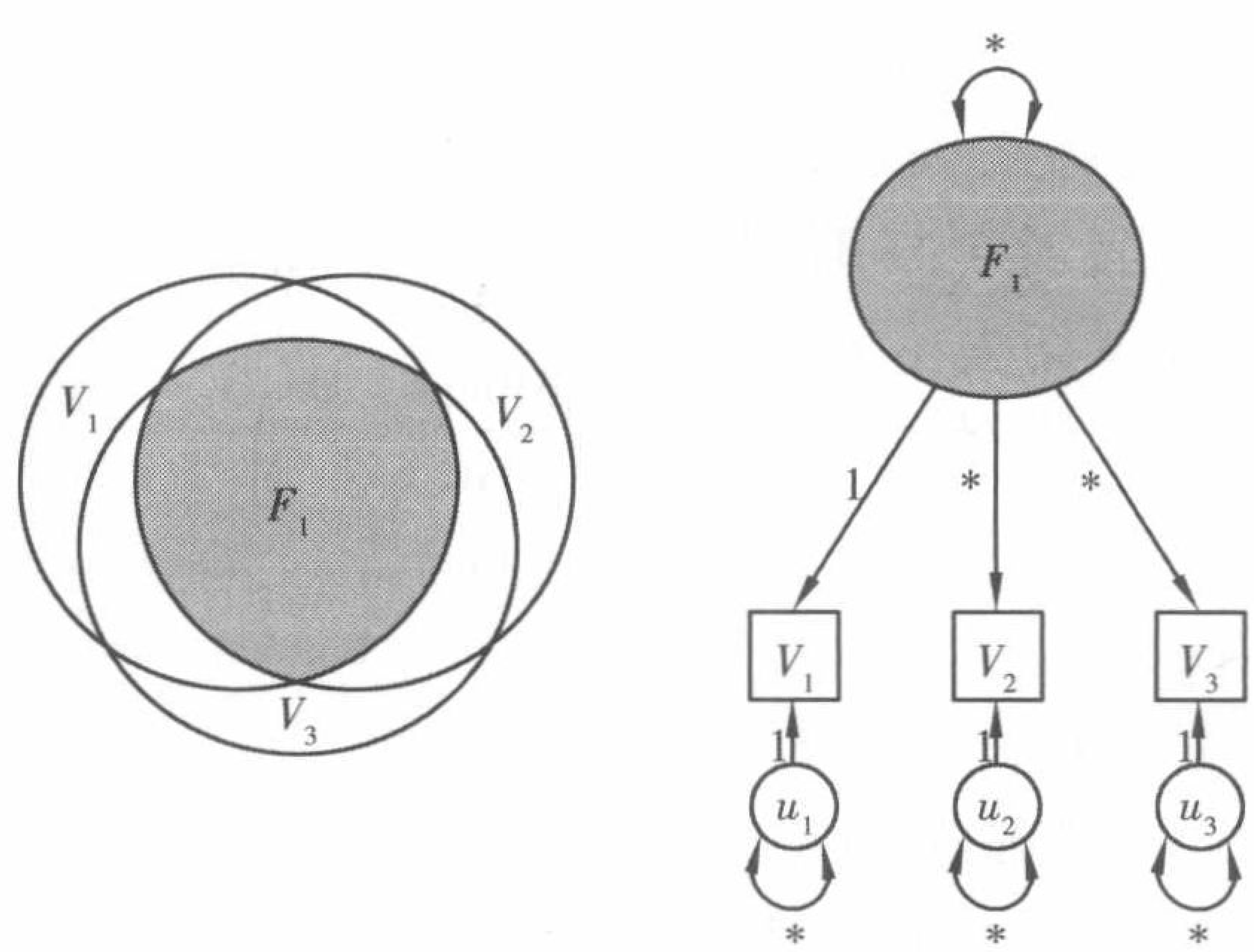



**图 1　将三个变量分数的变异性分解成共享和独有成分的 Venn 图和路径图**

在图 1 的右侧,是更为正规的这种变异性的分解图示。在这张路径图【Path Diagram】中,指标用方块表示。图中最大的标以 $F_1$ 的阴影圆圈与 Venn 图中的阴影区域相对应。这正是三个指标要测量和反映的它们共有的变异性的潜变量。潜变量和方块之间的方向路径表示潜变量是引起这些指标变异的原因这一假定。小圆圈表示每一个指标独有的变异,它与 Venn 图中的无阴影的部分相对应。它们代表着量度误差,因为它们是指标上计划之外的原因。Venn 图中的交叉线区域则未在路径图中有所表示。因为一个典型的量度模型的设定会对不相关的测量误差作出假定。如果这些关系的强度在统计上是非平凡的,那么一个令人满意的量度模型必须包括量度误差之间的协方差。

路径图的另外两个性质必须提一下。第一,注意每个圆圈上都连着一种急剧弯曲的双向箭头。这条路径表明一个方差【Variance】,并直观地阐明每个指标的变异性都已经分配给了潜变量或量度误差这一事实。第二,注意每一条路径都伴有一个 1 或星号。这些代表参数【Parameter】,反映路径强度的数值。那些以一个像 1 这样的数值表明的参数,反映一个两个变量之间的关联或一个方差的固定的假定。以一个星号表明的那些参数,表示参数的值虽未假定,但将会用观察数据对其作出估计。验证性因子分析特别感兴趣的那些参数的子集

则用潜变量和指标之间的路径——负荷相关联。这些值越大,指标中可以归结于潜变量的变异性的比例相对于量度误差而言就越大。

## 模型设定

验证性因子分析始于正式的量度模型的设定【Specification】。这就是说,研究者尝试以来源和说明指标之间关系的关联【Associations】路径的形式建立一组假设。这些假设参照指标的观察数据进行评价。为了简明地说明这一步骤,我们设想,有一个女性心理学家想评价一个新的测量学生的自我妨碍倾向(tendency to self-handicap)测量方法的效果。在这里,自我妨碍倾向是指一种学生在一些重要的评估活动,如测验和文稿演示之前的一种与预期目标相违的行为。她提出了这种倾向的两种独立的形式,并写出了每种倾向的三个题项。然后她将这个测量方法用于一个由 506 名大学生组成的样本。

研究者有关题项之间的关联的测量模型反映在图 2 中。两个潜变量 $F_1$ 和 $F_2$ 表示自我妨碍的两种假设的形式。注意 $V_1$ 到 $V_6$ 这 6 个指标,每一个都只受两个潜变量中的一个的影响。还要注意一下 $F_1$ 和 $F_2$ 之间弯曲的箭头。这种路径表示一种相关【Correlation】。因为研究者认为两种自我妨碍的形式之间是独立的,所以证明从她的数据估计得到的这一参数的值实际上是零,这一点对她来讲是很重要的。

在验证性因子分析的应用中,我们习惯将自己提出的测量模型与看似可信的备择模型进行比较。这些备择模型或较我们提出的模型简单,或与那些竞争的指标之间的关联理论解释一致。我们的假设研究者可能想把自己的模型与一个更为简单的只包含一个潜变量,或一个有一点问题,但不那么简单地包含两个有六个题项的、自我妨碍量度的、不独立的潜变量模型作比较。如果根据下面提出的标准,在绝对的意义上讲,如果研究者的模型不仅对数据作出了很好的解释,而且作出的解释也优于另两种备择方法作出的,那么她便有了她的量度因素效度【Validity】的强有力证据。

虽然验证性因子分析的模型设定有很大的灵活性,但是估计的数学要求还是对设定会有所限定。例如,一个有着如图 2 所示的每个指标同时受两个潜变量影响的模型,它尽管也相当于一个探索性因子分析【Exploratory Factor Analysis】模型,但却是无法估计的。一个没有一个负荷被固定为 1 的模型,若不改变模型的其他方面也是无法估计的。构成设定的这些约束的原则叫作识别。识别是一个具有挑战性的问题,幸而这个问题很少出现在如图 2 所示的一般性模型设定中。只要我们能坚持三个基本规则,那么得到的模型几乎总是可以被估计的。对于每一个潜变量而言,至少要有三个指标。每个潜变量的方差,或它上面的一个负荷必须固定在 1。此外,模型中不可以有比方差数和协方差数之和更多的未知的参数。在这个例子中,有 6 个方差,每一个对应于一个指标,有 15 个指标之间的协方差,总共有 21 个[这个数字可以用方程式 $p(p+1)/2$确定,式中的 $p$ 是指标数]。由计数图中的星号数可知,共有 13 个未知的参数。研究者的模型满足识别的标准,因而是可以被估计的。

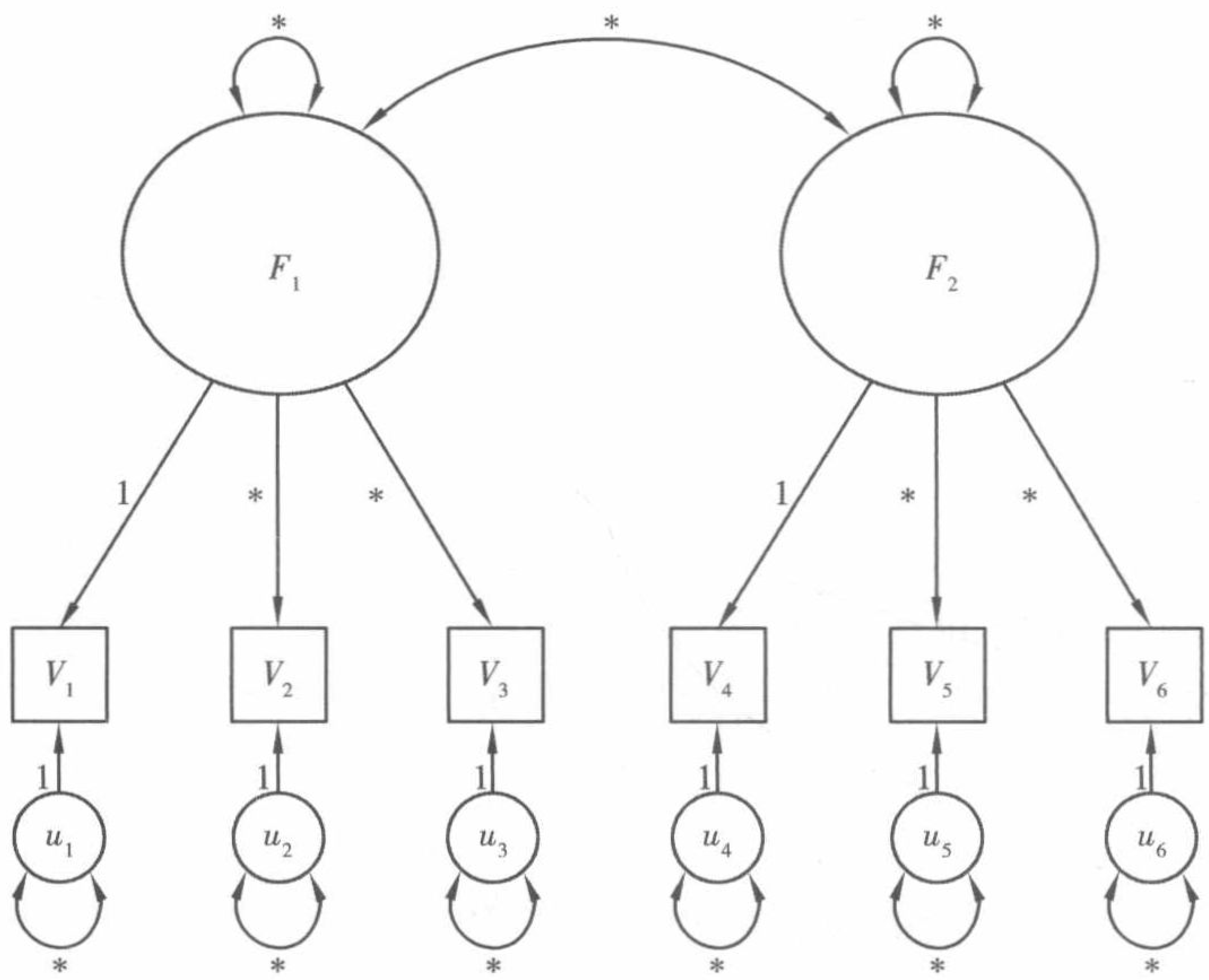



**图 2　双因子模型路径图**

## 参数估计

测量模型一旦设定,便可对它进行估计。这就是说,我们可以用指标上的数据得到未知参数的值。用于这一目的的估计量【Estimators】有若干个,但实际上首选的估计量总是最大似然估计【Maximum Likelihood Estimation】。最大似然估计的目的是确定未知参数最可能给予数据的值。这样的确定是通过一个替代未知参数的值,评估给予数据的参数的似然。然后试着通过这些值来改进似然达到的。这个过程是一个迭代过程。而当迭代再也没有可能对似然作出显著的改进时,评估过程便收敛了。收敛的参数值便是假设检验依据的那些参数值。没有暂时分配的未知参数的值,迭代过程便无法开始。解决这一问题的最简单的方法是先将 1 这个值分配给每一个未知的参数,然后再开始进行迭代。在第一次迭代时,这些初始值便会得到调整,有些向上,有些则向下,无论向上还是向下,第一次迭代一般都会导致已知数据的参数的似然显著提高。在第二或第三次迭代之后,似然的提高通常都比较小,一般在 10 次或不到 10 次迭代之后便会达到收敛。

这个例子的参数估计值列在了表 1 的上部。作为假设检验根据的参数估计值是非标准化【Unstandardized】的估计值。注意,在负载中,$V_1$, $F_1$和 $V_4$, $F_2$都是 1,与我们的模型设定中规定的一样。注意在“参数”这一标题下共有 15 个词条。如果我们排除那两个固定的负载,留下的 13 个词条,每一个都对应图 2 中的一个星号。

## 拟合度评估

在求出她收集的数据给出的她的模型最优参数估计之后,研究者必须要立即回答两个问题。第一个问题,模型是否令人满意地对数据作出了解释?第二个问题,如果模型对数据作出了令人满意的解释,那么参数估计是否符合预期的模式?第一个问题事关总括拟合(omnibus fit),而第二个问题则事关成分拟合(component fit)。

表 1 参数估计和总括拟合信息

| 参 数 | 参数估计值 | | |
|---|---|---|---|
| | 未标准化估计值 | 标准误差 | 临界比 |
| 负载 | | | |
| $V_1, F_1$ | 1.000[a] | | |
| $V_2, F_1$ | 1.236 | 0.257 | 4.804 |
| $V_3, F_1$ | 1.690 | 0.380 | 4.445 |
| $V_4, F_2$ | 1.000[a] | | |
| $V_5, F_2$ | 0.578 | 0.106 | 5.472 |
| $V_5, F_2$ | 0.694 | 0.127 | 5.472 |
| 方差 | | | |
| $F_1$ | 0.136 | 0.038 | 3.572 |
| $F_2$ | 0.870 | 0.175 | 4.690 |
| $U_1$ | 0.377 | 0.039 | 9.594 |
| $U_2$ | 0.881 | 0.073 | 12.034 |
| $U_3$ | 0.831 | 0.104 | 8.027 |
| $U_4$ | 0.683 | 0.158 | 4.326 |
| $U_5$ | 0.938 | 0.078 | 12.042 |
| $U_6$ | 1.347 | 0.112 | 12.024 |
| 协方差 | | | |
| $F_1, F_2$ | −0.004 | 0.026 | −1.61 |

| 模型 | 总括拟合指数 | | | | |
|---|---|---|---|---|---|
| | $\chi^2$ | $df$ | $p$ | CFI | RMSEA |
| 双因模型 | 13.820 | 8 | 0.087 | 0.976 | 0.038 |
| $F_1, F_2$ | 13.845 | 9 | 0.128 | 0.980 | 0.033 |
| $F_1, F_2$, add $U_1$, $U_6$ | 8.018 | 8 | 0.432 | 1.000 | 0.003 |
| 单因模型 | 100.149 | 9 | <0.001 | 0.622 | 0.142 |

注：CFI 为比较拟合指数；RMSEA 为近似误差均方根；a 为固定参数。

测量模型的总括拟合度是给定模型参数估计值隐含的一组数据与观察值近似的程度。在这里，数据一词并非 506 个参与调查的回答人对六个自我报告的题项的个别回应。相反，我们所谓的数据是观察的协方差矩阵的元素。我们的例子中观察的协方差矩阵列在了表 2 顶上的那一部分。该矩阵有 6 个方差——每一个对应于一个指标，以及 15 个协方差，代表六指标间的配对关联。

用于指标化拟合度的推论和描述性的统计量有很多。每一个都有合理的一面和解释。而在最基本的层面上，总括性拟合度的评估则是在比较中进行的。最为直截了当的比较，是观察到的协方差矩阵和来自数据的参数估计值给出的设定模型含有的协方差之间进行的比较。为了得到隐含的协方差矩阵，我们可以先将参数的估计和固定的值代入测量方程，进而求得一个协方差矩阵。我们的例子的隐含协方差矩阵列在了

表 2 中的第二部分。设定的模型恰好与这一理论矩阵相拟合。在这个意义上讲,这一矩阵与观察的协方差矩阵是匹配的,因此设定的模型给我们提供了观察数据令人满意的解释。

观察的和隐含的协方差矩阵之间的这种比较,是通过从观察矩阵中的对应元素减去隐含的协方差矩阵中的每一元素产生的残差矩阵完成的。该例子的残差矩阵列在了表 2 中的第三部分。这个矩阵包含了一些零,它表明没有其他的方差或协方差被解释。对角线上的零都是我们所预期的,因为指标中的所有方差都在模型中或由潜变量或由测量误差所解释。对角线之外的零表明研究者的模型充分解释了指标间的某些关联。在那些非零的残差中,有些是正的,有些是负的则表明模型低估了某些关联和高估了某些关联。尽管残差矩阵的信息量是很大的,但这仍然是不够的,特别是在模型需要重新设定时,它并未给我们提供正式判定的依据,即设定的模型在多大程度上对数据给予了满意的解释。

**表 2　观察的、隐含的和残差的协方差**

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | 观察协方差矩阵 | | | |
| $V_1$ | 0.513 | | | | | |
| $V_2$ | 0.169 | 1.090 | | | | |
| $V_3$ | 0.231 | 0.284 | 1.221 | | | |
| $V_4$ | 0.031 | −0.100 | −0.066 | 1.553 | | |
| $V_5$ | 0.035 | −0.031 | 0.052 | 0.503 | 1.229 | |
| $V_6$ | 0.106 | −0.017 | −0.020 | 0.604 | 0.351 | 1.767 |
| | | | 隐含协方差矩阵 | | | |
| $V_1$ | 0.513 | | | | | |
| $V_2$ | 0.169 | 1.090 | | | | |
| $V_3$ | 0.231 | 0.285 | 1.221 | | | |
| $V_4$ | −0.004 | −0.005 | −0.007 | 1.553 | | |
| $V_5$ | −0.002 | −0.003 | −0.004 | 0.503 | 1.229 | |
| $V_6$ | −0.003 | −0.004 | −0.005 | 0.604 | 0.350 | 1.767 |
| | | | 残差矩阵 | | | |
| $V_1$ | 0.000 | | | | | |
| $V_2$ | 0.000 | 0.000 | | | | |
| $V_3$ | 0.000 | −0.001 | 0.000 | | | |
| $V_4$ | 0.035 | −0.095 | −0.059 | 0.000 | | |
| $V_5$ | 0.037 | −0.028 | 0.056 | 0.000 | 0.000 | |
| $V_6$ | 0.109 | −0.013 | −0.015 | 0.000 | 0.001 | 0.000 |
| | | | 标准化残差矩阵 | | | |
| $V_1$ | 0.000 | | | | | |
| $V_2$ | 0.000 | 0.000 | | | | |
| $V_3$ | 0.000 | −0.001 | 0.000 | | | |
| $V_4$ | 0.039 | −0.073 | −0.043 | 0.000 | | |
| $V_5$ | 0.047 | −0.024 | 0.046 | 0.000 | 0.000 | |
| $V_6$ | 0.114 | −0.010 | −0.010 | 0.000 | 0.001 | 0.000 |

用于这一目的的总括性拟合指数有很多。其中最基本的是卡方统计量,这是拟合优度【Goodness of Fit】的推论检验的基础。在这个例子中,8 个自由度(自由度的计算方法是方差和协方差数 21,减去参数数 13)的评估卡方值是 13.82($p$ 值为0.09)表明观察的和隐含的协方差矩阵并无显著的不同。两个广泛使用的指数是比较拟合指数(CFI)和近似误差的均方根(RMSEA)。前者的值域为 0~1,而值大于 0.90 表明拟合度令人满意。在这个例子中,CFI 的值为 0.976。RMSEA 起源于 0,因而较小的值表明拟合得比较好。一般来讲,RMSEA 的点估计使用的是 90%的置信区间。因而在上限小于 0.08 则可表明拟合度是令人满意的。在这个例子中,上限是 0.071。在正式的典型的总括拟合度报告中,这一信息列在了表 1 底下部分的第一行。

拟合度评估的第二个层次涉及成分拟合,即参数估计值的数量和符号与预测相符合的程度。在这个例子中,我们期望因子的负荷是正的,且显著大于零。我们期望潜变量的方差也大于零。我们未敢建立有关量度误差的假设,不过较小的值表明,指标中归结于量度误差的方差也较小。最后,研究者假定自我妨碍的形式是独立的,因此我们预测 $F_1$ 和 $F_2$ 之间的协方差将与零相差无几。

现在我们再回到表 1 的顶部,看到来自数据的 13 个参数的估计值都伴随着一个标准误差。未标准化的参数估计值和它的标准误差的比率叫作临界比(critical ratio),它的分布与 $z$ 统计量相同。同样,值大于 1.96 表明估计值显著不同于零。在 13 个估计值中,只有 1 个不显著,而这正是我们假设的研究者认为应该为零的那个参数:$F_1$,$F_2$ 的协方差。

我们习惯于解释标准化的因子负荷,因为它们是用测量方程估计的,是标准化回归系数【Standardized Regression Coefficients】。经验法则告诉我们,值大于 0.30 表示潜变量对解释指标中的变异性有显著的贡献。所有六个标准化负荷都满足了这一标准。

我们研究者的模型对观察的数据在总括和成分拟合度两个层面都提供了令人满意的解释。为了进一步加强这一推论,她有可能将这一模型与我们早先描述的两个备择的模型进行比较。我们确定潜变量之间的相关与零没有什么不同,因此我们也由此而得知它也与初始的模型一样,这一参数将被设为零。而出于简约原因,我们更偏爱这一模型。第二个备择模型是一个单因子模型。这一模型的总括拟合指数列在了表 1 中的最后一行。注意,这一模型的总括拟合指数是高度显著的,且大大高于双因子模型。CFI 的值大大低于 0.90,且置信区间的上限和下限都超过了 0.08。因此在绝对和相对意义的两个方面,都给了我们研究者的模型以强有力的支持。

## 重设

在实际研究中,最初设定的测量模型并不会产生令人满意的总括拟合。在这种情况下,研究者可能想重新设定模型,以对数据作更为令人信服的解释。这个时候,我们的工作就从检验转向了假设的产生。严格地讲,统计检验不再是有效的了。尽管如此,在一个最初设定的模型的评估未曾给它支持时,重设却可能是十分有益的。

虽然在这个例子中,最初设定的模型情况良好,为了用它来说明重设问题,我们可以进一步问,拟合是否还有进一步改进的可能。我们已经知道,可以通过去掉 $F_1$ 和 $F_2$ 路径来简化模型。而这样一个模型的总括拟合指数已列在了表 1 中,它们支持了我们的假定。除了对模型作一个显而易见的调整之外,它并没有明确告诉我们总括拟合究竟得到了多大改进。在遇有这样的情况时,研究者求助于两种设定探索策略中的其中

一种。一种是几乎在所有用于验证性因子分析的计算机程序中都可得到的经验策略。另一种则是专注于残差矩阵的人工搜索策略。因为残差矩阵的元素都是协方差,我们无法立即对它作出解释,所以我们将它们标准化成如表2底部所列的标准化残差矩阵。这个矩阵的值是相关系数【Correlation Coefficients】,表示那些未被指标间的协方差解释的共变。这些系数中最大的是 $V_1$ 和 $V_6$ 之间的相关系数。最为简单的解释这种关联的方法是允许对应的量度误差,$U_1$ 和 $U_2$ 共变。在我们重新设定模型,将这一参数引进之后,标准化误差降到了0.024,而总括拟合度(列在了表1的底部),则接近了它们的最大值。如果没有初始设定的模型对这一个模型的支持,诸如这样的修改又如何能足以达到统计上的支持呢?然而我们必须牢记,因为拟合的诸如这样用观察数据的知识才达到的改进,是因样本而异的,因而在做先验和来自新数据的设定时,必须通过再设定模型的评估来验证。

## 渊源和最近的方向

验证性因子分析源于瑞典心理计量学家卡尔·约雷斯括格的工作。在1970年代初,约雷斯括格和他的合作者发明了一种算法,更重要的是,他也同时开发了分析协方差矩阵结构的计算机软件。他们的LISERL(linear structural relations,线性结构关系)模型使我们得以对协方差结构【Covariance Structure】模型进行正规的评估。而测量模型正是协方差结构模型的一个特例。

验证性因子分析的大多数进展都是广义协方差结构分析发展的副产品。这些发展包括计算机程序、某种允许用绘制路径图进行模型设定的方法、许多对卡方检验的不足有所改进的总括拟合指数、对涉及最大似然估计的数据没有硬性要求的估计方法等。验证性因子分析特有的进展则可为它在解决特有的研究问题上的能力所证明。这些例子包括涉及潜变量的交互效应【Interaction Effects】检验,潜增长曲线模型【Growth Curve Models】和状态特质模型(state-trait models)。

——Rick H. Hoyle

(沈崇麟译校)

## 参考文献

Hoyle, R. H. (2000). Confirmatory factor analysis. In H. E. A. Tinsely & S. D. Brown( Eds.). *Handbook of applied multivariate statistics and mathematical modeling* (pp.465-497). New York: Academic Press.

Jöreskog, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika*, *34*, 183-202.

Loehlin, J. C. (1998). *Latent variable models: An introduction to factor, path, and structural analysis* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum.

# 混杂(Confounding)

混杂这个词曾被用于至少三个不同的概念。在最早和最广泛的使用中,混杂是在估计因果效应中的一种偏倚源。这种偏倚有时又被非正式地叫作额外因子(或叫作混杂因素)和感兴趣的效应的混合效应。这种用法主导着非实验的研究,特别是流行病学和社会学的研究。第二种和较新的用法发端于统计学,混杂是分层量度或调整额外因子[一种被称作非坍缩性(noncollapsibility)的现象,或辛普森悖论(Simpson's paradox)效应]的

同义词。第三种用法源于实验设计【Experimental Design】的文献，混杂涉及一个特定设计的主效应和交互作用的不可分离性。这三种概念紧密相连，并非总是可彼此区分的。特别是混杂概念作为一种效应估计的偏倚和一种非坍缩性使用时，常常被等价处理，尽管它们是非等价的。在这里我们只对前一种概念进行介绍，这些概念的更为广泛的涉猎和比较请参见以下这些著作：Rothman & Greenland，1998，Chaps. 4，20；Greenland，Robins，& Pearl，1999；Pearl，2000；以及 Greenland & Brumback，2002。

## 作为效应估计中的偏倚的混杂

在有关混杂的经典讨论中，密尔（Mill，1843/1956，Chap. 10）在他的著作的第 10 章中明确地提出了"混杂效应"一词，虽然在这一著作的第三章中，密尔就提出了主要的问题，并承认弗朗西斯·培根是处理这些问题的先行者。他列出了旨在确定因果关系的实验【Experiment】的一个要求。"没有一种我们的确知道将会有效应的实验环境，允许它和那些其性质正是我们希望研究的代理混杂"（着重强调后加）。

在密尔的时代，实验（experiment）一词指在观察者控制的条件下实施的观察。这种用法仍常出现在日常英语中，而非一种对比试验的概念。尽管如此，密尔的要求建议，如果希望研究的代理不存在，那么就应该在我们的"实验"（基本上相当于一种没有控制的试验）结果，与我们期望的结果之间进行比较。如果这个结果与没有研究代理时我们期望的结果不同，那么密尔的要求确保期望之外的结果并非额外的"环境"（因子）造成的。然而，如果这些环境的确带来了期望之外的效果，且这个结果被错误地归结为研究代理的效应，那么这个错误就是额外效应和代理效应混杂（或混淆）的一种。

许多现代的文献遵循由密尔提出的同样的非正式概念化路数。这方面的术语现在更加专业了，用处理（treatment）一词专指由调查者操控的代理，而用暴露（exposure）来表示未操控的代理。超越密尔的主要发展在于，在没有研究暴露时，期望的结果现在几乎总是清楚地来自未处理的或未暴露的控制组的观察。无论一个人最终究竟在处理组还是控制组，只要自然或社会力量或个人的偏好会对其发生影响，一般混杂就会发生，不仅如此，这些力量或偏好也会对结果变量有影响。虽然这样的混杂通常都发生在观察研究中，但在处理【Treatment】的分配、操控和依从行为（compliance）中存在系统性不当的随机化实验中，也同样会发生混杂问题。有一个更为深入但存有争议的观点是，混杂（按照密尔原来的定义）在完美的随机化试验中，由于比较组之间的随机差别，同样也会发生（Fisher，1935；Rothman，1977）。

## 潜在结果模型

各种各样混杂模型曾被建议用于统计分析。而最接近密尔概念的一种模型是基于潜在结果，或因果效应的反事实【counterfactual】模型。假设我们的目的是要了解一种人口的健康状态量度（结果）究竟是怎么对一种干预（人口的处理）作出反应的。更为精确地说，假设我们的目的是对于目标总体 A，确定应用处理 $x_1$ 具有的效应，或在结果量度 $\mu$ 上的，相对于应用处理 $x_0$ 的相对效应。例如，同期群 A 可能是一个乳腺癌患者的同期群，处理 $x_1$ 可能是一种新的激素治疗法，$x_0$ 可能是一种安慰剂治疗法，而量度 $\mu$ 则可能是 5 年的存活概率。处理 $x_1$ 有时也叫作指数处理，而 $x_0$ 有时则也叫作控制或参照处理（它经常是一种标准

或安慰剂处理）。

潜在结果模型设想，在总体 A 中，如果应用了 $x_1$，那么 $\mu$ 将等于 $\mu_{A1}$；如果应用了 $x_0$，那么 $\mu$ 将等于 $\mu_{A0}$。$x_1$ 相对于 $x_0$ 的因果效应被定义为从 $\mu_{A0}$ 到 $\mu_{A1}$ 的变化，它可能用 $\mu_{A1}-\mu_{A0}$ 或 $\mu_{A1}/\mu_{A0}$ 来量度。如果 A 给予了处理 $x_1$，那么 $\mu$ 将等于 $\mu_{A1}$，且 $\mu_{A1}$ 是可观察的。但 $\mu_{A0}$ 则是不可观察的。然而，假设我们期望 $\mu_{A0}$ 等于 $\mu_{B0}$，而 $\mu_{B0}$ 是被施以处理 $x_0$ 的总体 B 的，观察或估计的结果 $\mu$ 的值。后一种总体有时叫作控制或参照总体。如果 $\mu_{A0}\neq\mu_{B0}$，那么我们就说有混杂存在，因为那时总体 A 和总体 B 之间必定存在着一些差别，正是这些差别（而非其他处理）在影响着 $\mu$。

如果混杂存在，由用 $\mu_{B0}$ 替代效应量度中的 $\mu_{A0}$ 求得的纯真（概约）的关联【Association】量度将不等于效应量度，因此说关联量度被混杂了。例如，如果 $\mu_{A0}\neq\mu_{B0}$，那么量度跨变量的处理和结果的关联的 $\mu_{A1}-\mu_{B0}$，对量度总体 A 中的处理 $x_1$ 的效应，$\mu_{A1}-\mu_{A0}$ 而言是被混杂了。因此，说一个关联量度 $\mu_{A1}-\mu_{B0}$ 对效应量度 $\mu_{A1}-\mu_{A0}$ 而言被混杂了，就是说，这两个量度是不相等的。

这种观点值得注意的第一点是，混杂取决于结果量度。例如，假设总体 A 和总体 B 在施以安慰剂处理 $x_0$ 时有着不同的 5 年存活率 $\mu$；那就是相当于假设 $\mu_{B0}\neq\mu_{A0}$，这样 5 年存活的实际处理效应 $\mu_{A1}-\mu_{A0}$，$\mu_{A1}-\mu_{B0}$ 已被混杂，然而在施以安慰剂处理时，两个总体的 10 年存活率，$v$ 仍然可能是相同的。也就是说，$v_{A0}$ 仍然可能等于 $v_{B0}$，于是 10 年存活率的实际处理效应，$v_{A1}-v_{B0}$ 是未被混杂的（一般我们可以期望 200 年存活的未被混杂，因为没有一种处理可以提高零岁以上病人的 200 年的存活概率）。

值得注意的第二点是，混杂取决于推论【Inference】的目标总体。前面的例子，用 A 作为目标，A 和 B 在用安慰剂治疗时有不同的 5 年存活率 $\mu_{A0}$ 和 $\mu_{B0}$，因此对总体 A 的处理效应 $\mu_{A1}-\mu_{A0}$，$\mu_{A1}-\mu_{B0}$ 是混杂的。也许有一位律师或伦理学家对激素治疗可能会对总体 B 产生什么效应感兴趣。若用 $\mu_{B1}$ 表示施以激素治疗的（未观察的）结果，那么 B 上的这一效应则由 $\mu_{B1}-\mu_{B0}$ 量度。用 $\mu_{A1}$ 替代未观察的 $\mu_{B1}$，产生 $\mu_{A1}-\mu_{B0}$。在这种情况中，当且仅当 $\mu_{A1}\neq\mu_{B1}$ 的时候，这一关联量度对 $\mu_{B1}-\mu_{B0}$（总体 B 中处理 $x_1$ 对 5 年存活的效应）而言才是混杂的。因此，同样的关联量度，$\mu_{A1}-\mu_{B0}$ 作为两个总体 A 和 B 上的处理效应，可能对二者都是不混杂的，或对其中一个是混杂的，或对二者都是混杂的；在用于其他一些目标总体的处理效应时，既可能是混杂的，也可能是不混杂的。

## 混杂因素（产生混杂的因素）

潜结果模型的第三个值得注意的方面是，它会在环境或可能会影响 $\mu$ 的协变量方面，在总体 A 和 B 之间引发一些不明确的（不平衡的）差异（Greenland & Robins, 1986）。问题很清楚，如果 $\mu_{A0}$ 和 $\mu_{B0}$ 不同，那么 A 和 B 在影响 $\mu$ 的那些因素方面也不同。观察到了这一点使有些研究者将混杂定义为在比较的总体之间存在着这样的协变量差异。然而，混杂只不过是这些协变量差异的结果。实际上，A 和 B 可能在影响 $\mu$ 的协变量方面有着很大的不同，然而却可以不会因此而产生混杂。换言之，A 和 B 之间协变量的差异是混杂的必要条件，但却不是充分条件。例如，协方差差异的混杂效应可能会彼此平衡而抵消，结果并没造成混杂。

假设总体 A 和总体 B 在某个协变量方面有差异，而这些差异导致用于我们感兴趣的效应量度的关联量度的混杂。于是造成混杂的协变量便叫作关联量度的混杂因子。在上面的例子中，对应于效应 $\mu_{A1}-\mu_{A0}$ 的 $\mu_{A1}-\mu_{A0}$ 被混杂了，造成混杂的那些因素（即导致

$\mu_{A0} \neq \mu_{B0}$的那些因素）就是混杂因素。由此可见，一个变量不可能是一个混杂因素，除非它能在处理组内影响结果参数 $\mu$，且在比较总体之间的分布有差异[具体例子请参见Yule，1903。不过他使用的是虚构关联（fictitious association）一词，而非混杂这个词]。这两个必要条件有时会一起提出来作为一个混杂的定义。

但是一些反例显示，这两个条件并非那些有两个以上水平的变量是一个混杂因素的充分条件。

## 混杂的预防

也许最为显而易见的避免混杂的方法是估计 $\mu_{A1}-\mu_{A0}$，以得到已知 $\mu_{B0}$ 等于 $\mu_{A0}$ 的参照总体 B。有时，我们说这样的总体和 A 就参照处理的结果而言，是可比的或是可互换的。实际上，这样的总体是难以或无法找到的。因此，调查者可能试图构建这样一个总体，或一个可互换的指数和一些参照总体。这些构建可看作是基于设计的控制混杂的方法。

也许对预防混杂而言更为有效的方法是了解而非限制一个因素。例如，性别失衡不会对只限于女性的研究造成混杂。然而，这种方法却有几个不足之处：过多地限制那些因素，可能会使使用的因素数降低到不可接受的水平，进而导致得到的结果缺乏普遍性。在混杂因素上匹配处理总体来克服这些不足，如果成功了，这种方法与限制法同样有效。例如，有着相同的女性比例的比较总体，就不会有因性别失衡而造成研究的混杂。

遗憾的是，观察损失数目的不同，会使由匹配产生的最初的协变量的平衡丢失。无论限制还是匹配，都不能预防（尽管它们可能减少）未限制的、未匹配的，或未量度的、协变量的失衡。相反，随机化给我们提供了一种处理研究设计解释不了的协变量造成的混杂的工具。然而必须强调，这种解决方法只是概率的，且在实际中要受到严格的约束。

随机化并非总是可行的或合乎伦理的。许多实际问题，如不同的损失数和不服从数会导致实际接受处理 $x_1$ 和 $x_0$ 的组的比较的混杂。一种多少还有一些争议的解决不服从问题的方法是意向-处理分析（intent-to-treat analysis）。它用处理的分配，而不是接受来定义比较组 A 和 B。然而混杂甚至也可能会影响意向处理分析，与普遍的意向错觉的偏倚相反，这种分析中的偏倚会夸大明显的处理效应（Robins，1998）。例如，在盲法（blinding）不足以阻止处理的提供者违反分配的协定时，分配就可能并非总是随机的。不仅如此，有时仅仅因为运气不好，随机化本身产生的分配结果，就有着严重的组间的协变量失衡（并因此而造成混杂）。在研究数量小的时候情况尤其如此（Fisher，1935；Rothman，1977）。区块（配对）随机化可以确保区块的因素不发生随机的失衡问题，但它无法保证未区块化的因素平衡。

## 混杂的调整

基于设计的方法常常是不可行或不足以预防混杂问题的。因此，需要做大量的工作来对混杂进行分析调整。除少数例外之外，这些方法都以比较总体中的观察的协变量的分布为根据。这些方法只有在某种程度上，即在有足够的混杂因素被精确地量度时，才能成功地控制混杂。此外还有许多其他在某个阶段采用参数模型的方法，而这些方法的成功需取决于模型对现实的正确程度。我们无法在这里对这些问题进行深入的讨论，但是若干基本点却值得在此一提。

最简单也是受到最广泛信任的调整法始于对混杂因素的分层。协变量不可能是造成协变量内在同质层的层内混杂的原因。

例如,性别失衡不可能造成只有女性组成的层内观察的混杂。更为广义地讲,一个和层内的处理无关联的协变量是不可能造成层内比较的混杂的。无论协变量是否用来作层的定义,情况都如此。因此,为了控制混杂,我们并不需要对所有的混杂因素分层。不仅如此,如果我们握有混杂因子之间的关系的精确背景资料,那么我们就可以用这些资料来确定一组协变量并足以用来进行调整了(Pearl, 2000)。然而,如果混杂因素的分层过于粗糙(例如因为类别的定义过于宽泛),分层则可能无法用调整变量对许多混杂进行调整。

当今,一种使用最为普遍的调整法是将那些存疑的混杂因素输入一个结果参数 $\mu$ 的模型。例如,令 $\mu$ 为感兴趣的结果变量 $Y$ 的均值(期望);令 $X$ 为感兴趣的处理变量;而令 $Z$ 为存疑的 $X—Y$ 关系的混杂因素。对 $Z$ 的调整经常通过拟合一个广义线性模型【Generalized Linear Model】$g(\mu)=\alpha+\beta x+\gamma z$ 或某些变体来做。式中的 $x$ 和 $z$ 是 $X$ 和 $Z$ 的特定值,而 $g(\mu)$ 则是一个严格的渐增函数,如在对数线性模型【Log-linear Modeling】中的自然对数 $\ln(\mu)$;或在 logistic 回归【Logistic Regression】中的 logit 函数中的 $\ln\{\mu/(1-\mu)\}$;而那样得到的 $\beta$ 的估计值,则被看作是 $X$ 的一个单位的增长在 $g(\mu)$ 上的效果的 $Z$ 的调整估计值。

基于模型的调整法一个常常被人引用的优点是,它可以进行多个变量的调整,且在细节上远胜于分层法。然而,如果拟合的模型不能与 $Y$ 在 $X$ 和 $Z$ 上的相依相适应,那么这样的基于模型的调整,就无法对 $Z$ 造成的混杂作调整。例如,假设 $Z$ 在 $X$ 的层内围绕零对称分布,而真的相依是 $g(\mu)=g(\alpha+\beta x+\gamma z^2)$,那么使用模型 $g(\mu)=g(\alpha+\beta x+\gamma z)$ 将对 $Z$ 无法,或只能作出很少的调整。在基于偏好分数的调整中,情况也同样如此。诸如这样的不足之处可以通过使用合理可行的模型,或对照数据仔细地检查每一拟合的模型来缩小或规避。

有关随机化试验中的协变量的调整问题曾经发生过论战。尽管费希尔(Fisher, 1935, p.49)声称随机化比较是"无偏倚的",但他也认为在这里使用的意义上,它们有可能被混杂。不过,若能注意到费希尔使用的无偏倚一词时,他指的是设计问题,并未打算把它用作一个给定的试验的分析指导这一点,问题便迎刃而解。试验一旦展开,且实际的处理分配也已完成,如果协变量和处理关联,那么未作调整的处理效应将会是有偏倚的,但这种偏倚是可以通过协变量的调整来去除的(Greenland & Robins, 1986; Rothman, 1977)。

最后请注意一个重要的问题:如果用于调整的一个变量并非一个混杂因素,那么混杂则有可能因这样的调整而引进(Pearl, 2000),如果这个变量同时受到研究中的处理和结果的影响,那么这种偏倚将会特别严重(Greenland, 2003)。

——Sander Greenland

(沈崇麟译校)

## 参考文献

Fisher, R. A. (1935). *The design of experiments*. Edinburgh, UK: Oliver & Boyd.

Greenland, S. (2003). Quantifying biases in causal models. *Epidemiology*, *14*, 300-306.

Greenland, S., & Brumback, B. A. (2002). An overview of relations among causal modelling methods. *International Journal of Epidemiology*, *31*, 1030-1037.

Greenland, S., & Robins, J. M. (1986). Identifiability, exchangeability, and epidemiological confounding. *International Journal of Epidemiology*, *15*, 413-419.

Greenland, S., Robins, J. M., & Pearl, J. (1999). Confounding and collapsibility in causal inference. *Statistical Science*, *14*, 29-46.

Mill, J. S. (1956). *A system of logic, ratiocinative and inductive*. London: Longmans, Green &

Company. (Original work published 1843)

Pearl, J. (2000). *Causality*. New York: Cambridge University Press.

Robins, J. M. (1998). Correction for non-compliance in equivalence trials. *Statistics in Medicine*, *17*, 269-302.

Rothman, K. J. (1977). Epidemiologic methods in clinical trials. *Cancer*, *39*, 1771-1775.

Rothman, K. J., & Greenland, S. (1998). *Modern epidemiology* (2nd ed.). Philadelphia: Lippincott.

Yule, G. U. (1903). Notes on the theory of association of attributes in statistics. *Biometrika*, *2*, 121-134.

## 结果效度(Consequential Validity)

结果效度【Validity】关注社会影响的评估。《教育与心理测验标准》(*Standards for Educational and Psychological Testing*)(American Educational Research Association, 1999)把效度看作一种统一的概念。这一概念旨在测量所有累积的证据在多大程度上,就建议的目的而言,支持了测试分数的预期解释。因此,结果效度是证实意料的和意外的测试结果的解释和使用的那样一种类型的效度证据(Messick, 1989, 1995)。

涉及需要确定一个工具的结果效度的有害和意外的结果的例子有以下几种:

1.在将一个单一的测试分数用于决策的时候,高风险的测试可能给目前和将来的教育和就业机会以负面的影响。

2.对某些子群体,如残疾人或种族/民族的少数族裔进行测试时,由于不平等的机会,或在使用的书面或口头语言方面的不恰当的测试要求可能会导致低及格率。

3.那些会增加被测试者不愿寻找工作,或不愿完成要求的测试的职业考试会导致劳动力缺乏多样性。

4.教师可能会缩减课程,只讲授那些要测验的课程。

梅西克(Messick, 1995)建议研究者要区别源于个体和群体差异的效度描述的不利结果(adverse consequences),与源于诸如建构【Construct】代表名额不足(underrepresentation)和不相干建构(construct-irrelevant)方差的无效度测试的不利结果。如果差异与后者有关,那么测试无效度则显示了一个需要在验证过程进行调查的量度问题。然而,如果差异源于有效的群体差异,那么不利结果则显示了社会政策的问题。研究者的责任在于保证不会因为在会影响被试者展示自己能力的建构(如果建构确实存在)中丢失了什么东西而导致低分。此外,测试过程中那些不相干的要求会妨碍被试者证明自己的才能,从而会导致低分。

《教育与心理测验标准》也强调必须要把测试结果使用的那些与建构测试效度有关的方面和社会政策区别开来。标准1.24(Standard 1.24)阐述了结果效度问题:

"在意外结果源于测试的使用时,应当调查一下这样的结果是否源于对特征的测试敏感性,而非那些打算要评估的特征,或测试没有完全代表预期的建构"(American Educational Research Association, 1999, p.23)。

群体差异本身并不会对提出的解释的有效性产生怀疑。不过它们的确会增加对可能与量度效度有关的竞争假设进行调查的需要。

——Donna M. Mertens

(沈崇麟译校)

### 参考文献

American Educational Research Association, American Psychological Association, & National Council on Educational Measurement. (1999). *Standards for*

*educational and psychological testing*. Washington, DC：American Psychological Association.
Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13-103). New York：American Council on Education.
Messick, S. (1995). Validity of psychological assessment. *American Psychologist*, *50* (9), 741-749.

## 一致性(Consistency)

参见参数估计【Parameter Estimation】。

## 常数(Constant)

常数这一词无非是指那些不变的事物。在统计学中,响应一般都被看成随机变量【Random Variable】,它的大致含义是,响应是不能确定无疑地预测的。例如,按照某种特别的规定进食,普通的成年人将会减少多少体重? 个人对某种新药的反应如何? 如果我们随机地挑选某个人,他或她是否赞成死刑?

虽然在某种层次上,常数和随机变量之间的差别是清楚的,但在许多时候两者的区别常常是模糊不清的。例如,让我们来看一下总体均值$\mu$。这就是说,如果个体是可以量度的,$\mu$就是一特定研究中我们感兴趣的所有的个体的平均值。所谓统计问题的频率论路数,就是把$\mu$看成一个常数。它是某个固定的,但却是未知的值。然而,一种备择的,来自贝叶斯统计法的观点却不把$\mu$看作一个常数,而看作一个有着某种分布的数量。这种分布可能反映事先确信$\mu$有某种特定的值的可能性。作为另一种例子,$p$可能是一个人在回答他或她的婚姻是否幸福时回答是的概率。在一个特定的时间,我们可以把$p$看作在所有的已婚夫妇中是固定的。与此同时,无论在调查者事先确信,还是在随时间变化而变化的意义上,我们也可以把$p$看作是一个随机变量。

另一常数这一概念起着重要作用的一般语境,与分析时作的假定有关。这时,我们常常假定数据的某种性质是常数,以使技术问题简化。这方面最为著名的例子也许是同方差性【Homoskedasticity】。这是指一个我们常作的假定:个体群之中的方差【Variance】是一个常数。在回归【Regression】中,常数方差意味着在基于某个变量$X$预测$Y$时,给定$X$的$Y$的(条件)方差是不变的。正因如此,例如,如果$X$是家庭中的攻击性(aggression in the home),而$Y$是一个认知函数量度,那么常方差的意思就是,在有攻击性分数,如$X=10$的家庭中,与在有$X=15$、$X=20$或任何其他可能选择的$X$的值的家庭中,$Y$的方差都相同。

——Rand R. Wilcox

(沈崇麟译校)

### 参考文献

Conover, W. J. (1980). *Practical nonparametric statistics*. New York：John Wiley.
Doksum, K. A., & Sievers, G. L. (1976). Plotting with confidence：Graphical comparisons of two populations. *Biometrika*, *63*, 421-434.
Wilcox, R. R. (2003). *Applying contemporary statistical techniques*. San Diego：Academic Press.

# 持续比较法(Constant Comparison)

持续比较法是一个数据分析的过程,从数据分析中得到的每个解释和发现都要和已有的发现作对比。它与定性研究【Qualitative Research】的关联比跟定量研究【Quantitative Research】的关联更为密切。一般来说,它与扎根理论【Grounded Theory】的数据分析方法相关。在该方法中,格拉泽和斯特劳斯(Glaser & Strauss,1967)把它称作"定性分析中的持续比较法"。定性数据和定量数据都可以拿来作持续比较,但是那些数据的这种分析总是定性的。

通常每次比较都被称为一次迭代,而且一般都与归纳【Inductive】推理而非演绎推理更相关(Silverman,1993)。不过,在持续比较法的每次迭代中常常会发生"假设—演绎"的推理。持续比较法常关系具体的【Idiographic】哲学和研究方法,而非一般的哲学和方法。经常采用持续比较法的方法包括民族志【Ethnography】、现象学【Phenomenology】、符号互动论【Symbolic Interactionism】以及常人方法论【Ethnomethodology】。持续比较法有助于提高研究的效度。

组织内的领导现象研究便可视为持续比较法的一个显而易见的范例。这种方法可以采用观察和文献资料支持的访谈数据。对那些数据的初步分析可能会涉及对访谈录音整理稿进行编码【Coding】,以确定变量或分类,也即那些似乎已在现象中表现出来的东西。在对初步访谈进行分析之后可能会出现许多类,也许还可以指出这些类之间的关系。在接下来的每一次访谈中,每个出现的类都拿来和已有的类作比较,以确定这个新出现的类是离散的类别还是已有的类的一个特性,还是在一个更高的抽象层级中的某一类的代表。

举例来说,卡恩(Kan,2002)采用全扎根理论法来研究一个公立医院里护理工作的领导。她花了 14 个月的时间进行追踪访谈和观察,并从这些访谈和观察的持续比较中发现了许多低阶的类。而且卡恩在这段时间里还进行了问卷调查,以测量要研究的管理者所展现出来的领导力。把问卷结果与观察和访谈中得到的类相比较,就得到了两个高阶的类,分别叫作多重现实(multiple realities)和压制性的领导(repressing leadership)。

这种在整个数据分析中一直不间断的,或者说持续的比较,贯穿整个数据分析,连续不断地进行,直到弄清楚所有类别的特征和所有类别之间的关系为止。在卡恩(Kan,2002)这个例子中,对某些因素的低信度进行的探讨和对问卷作出的书面评论为我们提供了洞见,凸显了高阶类别的特点,并最终确证了"认同悖论(identifying paradox)"这个基本的社会过程。

——Ken W. Parry

(杨可译 沈崇麟校)

## 参考文献

Glaser, B., & Strauss, A. (1967). *The discovery of grounded theory*. Chicago: Aldine.

Kan, M. (2002). Reinterpreting the multifactor leadership questionnaire. In K. W. Parry & J. R. Meindl (Eds.), *Grounding leadership theory and research: Issues, perspectives and methods* (pp. 159-173). Greenwich, CT: Information Age Publishing.

Silverman, D. (1993). *Interpreting qualitative data: Methods for analysing talk, text, and interaction*. London: Sage.

## 构念(Construct)

构念是社会科学家使用的概念,他们借此来解释某一现象的经验资料,或者将某个领域不可观察、不可测量的因素加以概念化,以构建一种理论。具体来说,它可以表示为:(a)一个不可直接观察或测量的理念或概念(也即某个理论的某一方面);(b)从经验资料中推论或建构出来的一个概念,用以从整体上解释一个现象;(c)在其他可观测的经验概念或结果中显现出来的现象或概念的抽象的(而非操作性的)定义。例如,根据某些智力理论的说法,逻辑推理就是一种构念,它代表了智力的一个方面。逻辑推理不能直接观察;它是从诸如解决问题、批判性思维的能力等可观测的指标中推论出来的。(学习)动力也是一个构念,它可以从有关学习的经验资料中推出来;它促进不同数据的整合,以建构一个有关学习的理论。数学能力也是对一种行为的抽象定义;它的存在是从诸如数学成绩、数学焦虑(mathematical anxiety),以及某人曾经上过的数学必修、选修课等表现中推论出来的。

在构建个性和社会智力理论时,心理学家使用了普通人提供的语言和非语言的判断资料,以及多维度量【Multidmensional Scaling】来识别和确认各种构念(Jackson, 2002)。这种方法也同样成功地应用于心理病理学的临床判断研究(Jackson, 2002)。还有学者(Digma, 1990)用因子分析【Factor Analysis】来研究亲和性、意志力、情绪稳定性以及理解力(或开放性)等人格维度。这些构念(也即因子)是从测量人格的许多工具的相关【Correlation】或协方差【Covariance】矩阵中提取出来的。利用结构方程模型等先进的统计推论技术,卡罗尔(Carroll, 2002)说明了从这五个因素中还可以进一步提取出高阶因子来解释个人的人格差异,并协助诊断和治疗人格缺陷。在教育评估中,按照1999年的《教育与心理测验标准》(*Standards for Educational and Psychological Testing*)的有关条款,构念是一种需要收集起来用以支持某种评估工具的有效性宣称的证据。梅西克(Messick, 1992)还声称:"分数解释的构念效度可以支持所有以分数为根据的推论,不仅与那些涉及阐释意义的推论有关,而且还与那些专门以测试分数为根据的应用性决策及行动的、关系到内容和标准的推论相关"(p.1493)。对类似全国教育进展评估考试(the National Assessment of Educational Progress, NAEP)等基于标准的评估而言,测量构念是由内容标准和对学生理解和表现的不同方面的不同侧重来塑造的。

一个构念若要对研究社会现象有用,(a)它必须界定清晰;(b)必须具体说明它与相同研究领域内相似或不相似的概念或构念的关系。对一个构念的定义,以及对它与该研究领域其他因素间关系的阐释有时被合起来称作"构念理论",或者"基则网(nomological net)"(Mueller, 1986)。尽管构念可能是抽象的、不可观察和不可测量的,但它的表现或者效果必须是可以让社会科学家用某种方式观察和测量的,以便对它进行研究、构造理论,或者检验有关构念的假设。

——Chao-Ying Joanne Peng

(杨可译　沈崇麟校)

*也可参见构念效度【Construct Validity】。

### 参考文献

American Educational Research Association, American Psychological Association, & National Council on

Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.

Carroll, J. B. (2002). The five-factor personality model: How complete and satisfactory is it? In H. I. Braun, D. N.Jackson, & D. E. Wiley (Eds.), *The role of constructs in psychological and educational measurement* (pp. 97-126). Mahwah, NJ: Lawrence Erlbaum.

Digman, J. M. (1990). Personality structure: Emergence of the five-factor model. *Annual Review of Psychology*, *41*, 417-440.

Jackson, D. N. (2002). The constructs in people's heads. In H. I. Braun, D. N. Jackson, & D. E. Wiley (Eds.), *The role of constructs in psychological and educational measurement* (pp. 3-18). Mahwah, NJ: Lawrence Erlbaum.

Messick, S. (1992). Validity of test interpretation and use. In M. C. Alkin (Ed.), *Encyclopedia of educational research* (6th ed., p. 1493). New York: Macmillan.

Mueller, D. J. (1986). *Measuring social attitudes: A handbook for researchers and practitioners*. New York: Teachers College, Columbia University.

## 构念效度(Construct Validity)

在教育评估和社会科学的测量中,构念效度是在检验验证的背景下来定义的。检验验证是一个借助数据收集,给根据测验分数作出的推论建立可信性的过程——既包括关于使用测验分数(例如在制订决策时)的推论,也包括对数据(意义)的阐释的推论。在1985年的《教育与心理测验标准》(*Standards for Educational and Psychological Testing*)(American Educational Research Association, 1985)提出之前,效度是按照行为学家的传统来概念化的,有三个主要的不同的大类(即内容效度、效标效度【Criterion Related】和构念效度)。构念效度是指一个测验所体现出来的测量理论构念或某种特质的程度。像创造力、诚实、智力等构念都是社会科学家用来解释关于某种现象的经验资料或构建理论的概念。为了收集效度的证据,研究者必须首先定义要研究的构念,然后产生有关构念的结构和它与其他构念以及任何其他人类特性之间的关系的有理论根据的假设("有效性假设")。最后用经验数据来支持或否定这些有效性假设。一个完整的验证需要检验许多有效性假设。因此,检验验证是一个累积的过程。

人们提出了许多可以支持构念效度的技术,包括因子分析【Factor Analysis】、收敛和判别验证(convergent and discriminant validation)[Campbell, & Fiske, 1959年提出的多元特质——多种方法矩阵(multitraitmultimethod matrix)]、成长变化考查(例如,智力测验的年龄差异)、内部一致性、实验干预、已知群体差异(known-group differences)和与其他量的相关,或利用那些假设与研究的构念有关联的特质或行为领域的检验(Anastasi & Urbina, 1997; Cronbach & Meehl, 1955)。

洛文杰(Loevinger, 1957)提出,构念效度包括内容效度和效标关联效度。这个观点和其他一些观点一起促使1985年版和最新版的《教育与心理测验标准》放弃了三种效度的概念,取而代之的效度被构建为"一个一体多面的概念"(Messick, 1989)。根据这一概念,我们必须收集不同类型的证据来支持有关测试分数的意义以及正确使用方法的推论。必须强调一点,有效的是有关测试分数的解释和使用方法的推论,而非测试本身。因此,构念效度就是验证过程——一个收集证据和作出推论的系列——的中心。

像美国教育研究协会(the American Educational Research Association, AERA)、美国心理学协会(the American Psychological Association, APA)以及美国国家教育测试委员会(the National Council on Measurement in Education, NCME)这样的专业机构,建议应考虑不同的效度证据来源,包括而不限于下列内容:

- 测试标准的证据;
- 基于与其他变量的关系的证据(例如,测试分数与外部变量的相关性);
- 基于测试内容的证据;
- 收敛和判别的证据;
- 基于内部结构的证据(即测试项目和测试构成与一个理论构建相符),以及普遍性(测试标准可不对新环境的效度进行额外研究,就可普遍用于新环境的程度)。

梅西克(Messick, 1989, 1992)提出了表1中的矩阵,简要总结了他对测试效度模型概念化的关键概念。

表1 效度各个方面的渐进矩阵

| | 测试解释 | 测试使用 |
|---|---|---|
| 证据基础 | 构念效度(CV) | CV+关联性/用途(R/U) |
| 结果基础 | CV+价值意涵(VI) | CV+VI+R/U+社会后果 |

需要注意的是,研究设计领域在使用构念效度时,定义略有不同。按其定义,要关注的是处理/操作的精确性。处理是否对预期变量的操作有所影响?这个处理是否符合自变量所代表的构念的理论定义?就这种情况而言,该术语有时被称为因果构念效度(Kirk, 1995)。

——Chao-Ying Joanne Peng

Daniel J. Mueller

(杨可译 沈崇麟校)

*也可参见构念【Construct】、效度【Validity】。

## 参考文献


American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1985). *Standards for educational and psychological testing.* Washington, DC: American Psychological Association.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing.* Washington, DC: American Psychological Association.

Anastasi, A., & Urbina, S. (1997). *Psychological testing* (7th ed.). New York: Prentice Hall.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod validity. *American Psychologist*, *15*, 546-553.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*, 281-302.

Kirk, R. E. (1995). *Experimental design: Procedures for the behavioral sciences* (3rd ed.). Belmont, CA: Brooks/Cole.

Loevinger, J. (1957). Objective tests as instruments of psychological theory. *Psychological Reports*, *3* (Suppl. 9), 635-694.

Messick, S. (1989). Meaning and values in test validation: The science and ethics of assessment. *Educational Researcher*, *18*(2), 5-11.

Messick, S. (1992). Validity of test interpretation and use. In M. C. Alkin (Ed.), *Encyclopedia of educational research* (6th ed., p. 1487-1495). New York: Macmillan.

# 社会建构论(Constructionism, Social)

首先,社会建构论是对知识生产实践的一种表述——既有科学解释,也有其他解释。在这个层面上,建构主义理论提供了科学的知识生产的取向,一个与经验主义传统差异甚大的观点。同时,社会建构论也包含了人体机能理论的一些成分;在这个层面上,它为个人心理过程的传统观点提供了一种替代的观点。建构论的前提也已被延展到许多实践领域,在诸如治疗、组织管理、教育等领域开启了新的起点(Gergen, 1994, 1999)。建构论与社会科学的关系更非一般,正是因为有它们,才使许多新式研究方法得以在社会科学中盛行。

## 社会建构论的假定

社会建构论不能被还原为一套固定的原则,而应当被视为一段不断展开的关于知识的本质和对世界的理解的对话。不过,在被认定为建构论的作品中可以找到一些典型的主题。一开始,建构论通常假定我们对世界的描述——科学的描述和其他的描述——不是由存在是什么这种原则性的方式来述说和决定的。相反,我们用来理解世界的术语一般都被认为是社会的人工产物,是历史上人与人之间互动的产物。因此,在某个文化内,一个既定的理解形式盛行,在根本上并不取决于这里讨论的观点的经验效度,而是由社会过程(例如,沟通、协议、教派冲突、修辞技巧)的兴衰枯荣决定的。这个推理脉络完全不去贬低各种形式的文化理解的重要性,无论是科学的还是其他的。人们对世界和自身的建构,对于更为广泛的文化实践来说是最基本的——辩护、维持和变换各种行为方式。此外,不同的制造意义的社群也会给人类可资利用的资源带来不同的贡献——无论是"医学治疗"、"道德上的可理解性"、法律制度,还是"活着的理由"。无论如何,建构论确实对所有群体(也包括科学)超出其边界宣称的"真理"的正当性都提出了挑战。在一个传统之内的正确、真实和善行,在另一个传统之内可能就不是,因而不存在传统之间进行判断的标准,也根本不存在脱离传统、自己的价值观、目标和生活方式的标准。

## 社会建构和社会科学

这种综合发展所喜好的社会建构论观点开始成了社会科学中传统的经验主义描述的替代。在这个替代过程中,可以划分为两个阶段,即解构和重构。在前一个阶段中,科学理性,以及已在经验上证明正当的知识观的主体都受到了质疑。这个工作基本上表现为对早期反基础主义论述的细化和延伸,在当今,则主要是受了文学和批判领域中其他发展的启发。于是出现了大量文献,对真理宣称、经验假设检验、普遍理性、人体机能的规律、科学的价值中立、西方科学实践的输出等的合理性以及意识形态的含义提出了质疑。

只沉浸在这种文献中会得出一个结论,即社会建构论的目标是虚无主义的。不过,正如许多人相信的一样,解构的过程只是重构事业的一个必需的过程。重构阶段的主要目标就是科学提问的方式,在建构论的观点指导下可以更有效地为其所属的社会服务。这种新的敏感带来了一些值得注意的重大的发展。首先,建构论的思想非常重视理论创新,而非"照世界的样子把它画下

来”,其吸引力就在于创造可理解的问题,它也许会有助于我们建设新的未来。合作认知理论、半机械人政治(cyborg politics)、行动者网络都可作为例子。其次,建构论启发了文化研究的许多工作,对日常生活实践和产品进行了批判性、启蒙性的检视。最后,特别是在治疗、组织变革和教育中,建构论的想法也形成了许多新的实践。许多学者还发现,建构论在对知识的学科疆界提出挑战的同时,带来了范围广泛的对话。于是培养出了新的兴趣领域,比如,把神学和建构论、文学理论和社会运动、人格研究和伦理学理论联系在了一起。

## 社会建构和研究方法

尽管许多建构论的作品都是对社会科学传统的经验主义方法的批评,这些批评却并不致命。建构论思想绝没有要求将一种研究方法作为另一种研究方法的对立面;每种方法都有其建构世界的方式。因此,尽管传统的经验主义方法可能被视为有局限的,意识形态上也有问题,但它们也有重要的用途。无论如何,在方法论领域中建构主义思想最重要之处就在于激起对新的研究方法的讨论。尽管人们倾向于认为这些新方法是“定性的方法”(Denzin & Lincoln, 2000),建构论者并没有受制于传统上的所谓定性/定量,并认为前者较后者初级和低级的看法。大多数定性研究的目标、价值观和政治都和那些内在定量的研究不同。如今,建构论对强调以下内容的研究方法尤为青睐:

- 价值反思:这些研究方法对谁有利,让谁受损?研究对象是否会受到研究的剥削,或者是否仅被当作一个物体?
- 研究对象的声音:是否能让我们听到研究对象的声音?该研究方法是让研究对象的声音合法化,还是在研究程序中就去掉了他们的声音?
- 合作参与:研究对象是否能加入研究者中,一起研究知识?他们是否能分享研究结果或从中受益?
- 多重立场:研究是体现了多种观点和价值观,还是只有一种观点在主导?
- 表象主义的创造性:对研究的表现是仅限于正式写作,还是可以找出更亲民的、丰富的引人注目的表现方式?就研究重点而言,建构论的假设在参与行动研究【Participatory Action Research】(Reason & Bradbury, 2000)、话语分析(Wetherell, Taylor, & Yates, 2001)、叙事研究(如 Josselsyn, 1996)、参与式民族志、文学方法和述行方法(performative approaches)中的表现特别明显。

——Kennet J. Gergen

(杨可译　沈崇麟校)

## 参考文献


Denzin, N. K., & Lincoln, Y. S. (Eds.). (2000). *Handbook of qualitative research* (2nd ed.). Thousand Oaks, CA: Sage.

Ellis, C., & Bochner, A. P. (Eds.). (1996). *Composing ethnography*. Walnut Creek, CA: AltaMira.

Gergen, K. J. (1994). *Realities and relationships*. Cambridge, MA: Harvard University Press.

Gergen, K. J. (1999). *An invitation to social construction*. London: Sage.

Josselsyn, R. (Ed.). (1996). *Ethics and process in the narrative study of lives*. Thousand Oaks, CA: Sage.

Reason, P., & Bradbury, H. (Eds.). (2000). *Handbook of action research, participative inquiry and practice*. Sage: London.

Wetherell, M., Taylor, S., & Yates, S. J. (Eds.). (2001). *Discourse theory and practice*. London: Sage.

## 建构论(Constructivism)

尽管建构论的各种形式遍及哲学和科学,但在心理学中它是作为一种学习理论出现的。它基于皮亚杰(Piaget,1929, 1937/1955)对儿童的研究,该研究表明,儿童从他们的经验中去积极建构其自身的知识,而不只是吸收老师们所讲的。布鲁纳(Bruner,1960,1966)后来采用了这个理论并将其应用于教育。他建议在设计教学方式时应鼓励孩子思索和使用自己已有的知识和能力。这样,他们就可以通过为自己解决问题来理解新的概念。

不过,建构论也被更为广泛地用以指称心理学中那些认为人人都积极参与了自己主观世界建构的理论和方法。与之对立的是那种认为物体和事件都有基本属性和普遍意涵的观点,在这种观点看来,感知的关键在于将对世界的真实呈现内在化。格拉瑟斯菲尔德(Von Glasersfeld,1984)创造了激进的建构论这个术语,他提出,我们在世界上所有的知识都是建构出来的,而非通过感知得来的。因此,我们都是自己的主体现实的建构者,因为我们每个人都亲身栖居在此世界中。举例来说,曾经目击了同一事件的人,对该事件的回忆和描述不尽相同,这非常常见。激进的建构主义会说这不是因为感知和记忆不可靠,而是因为每个人都以不同的方式建构了这个事件。激进的建构主义从不在我们自己的建构之外宣称什么客观现实的存在或本质。

乔治·凯利的立场与此相仿,他是个人建构心理学(PCP)的创始人。凯利(Kelly,1955)认为,我们每个人都发展了一套多维度的意义系统,称为构念。我们用这些构念来感知这个世界,而我们的行为则可根据我们自己对世界的特有的解释来理解。作为一个临床心理学家凯利还设计了一个非常灵活的办法来考察就诊者的构念,他将这个办法称为凯利方格技术【Repertory Grid Technique】。此后,这个办法一直被当作研究工具,用于临床领域之外的许多环境。

建构论与社会建构论有一些相同的基本假定,但对何种程度上个人可以被视为控制建构过程的主体,以及在何种程度上我们的建构是社会力量的产物,这两个问题的理解还不尽相同。基亚里和努佐(Chiari & Nuzzo,1996)试图为区别各种建构论提供一个基础,建议对认识论的建构论和诠释学的建构论加以区分。在诠释学的建构论(也包括社会建构论)看来,现实是由语言和话语组成的。认识论的建构论(包括了个人建构心理学,PCP)则强调通过对经验的整理和组织来建构我们的个人世界。

——Vivien Burr

(杨可译　沈崇麟校)

### 参考文献

Bruner, J. (1960). *The process of education*. Cambridge, MA: Harvard University Press.

Bruner, J. (1966). *Toward a theory of instruction*. Cambridge, MA: Harvard University Press.

Chiari, G., & Nuzzo, M. L. (1996). Psychological constructivisms: A metatheoretical differentiation. *Journal of Constructivist Psychology*, *9*, 163-184.

Kelly, G. (1955). *The psychology of personal constructs*. New York: Norton.

Piaget, J. (1929). *The child's conception of the world*. New York: Harcourt Brace Jovanovich.

Piaget, J. (1955). *The construction of reality in the child*. New York: Routledge Kegan Paul. (Original work published 1937)

von Glasersfeld, E. (1984). An introduction to radical constructivism. In P. Watzlawick (Ed.), *The invented reality*. New York: Norton.

# 内容分析(Content Analysis)

内容分析最早有一种定义是:“我们所了解的内容分析这种技术……试图用系统的、定量化的方式来对给定的一长段话语的意义进行概括。”(Kaplan,1943,p.230)“内容分析是对交流的外显内容进行客观、系统、定量化描述的研究技术。”(Berelson,1954,p.489)后来的定义去掉了对定量化的要求。研究重点转为“推论”“客观”和“系统化”。在早期的一本教科书中,霍尔斯蒂写道:“内容分析可以是通过客观、系统地识别信息的指定特征,并从中作出推论的任何技术……我们的定义不包括定量化。”(Holsti,1969,p.14)此后,对内容分析技术发展作出贡献的其他学者也都重复了这一论点(如 Krippendorf, 1980)。无论是定量的推论,还是简单的推论,内容分析都将文本研究当作一种科学的而非简单的解释活动。

## 一个不同方法的集合

在内容分析这个共同的标签之下并非只有一种技术;相反,这个标签下是一个分析文本,或者更广义地说,分析各种信息的不同方法的集合——从句法分析的最简单的数字数到主题分析、指涉分析(referential analysis)、介词分析(prepositional analysis)(Krippendorf,1980,pp.60-63)。在这个列表后面,我们还必须加上最新的基于故事文法与社会关系有关的研究方法。

## 上下文中的字数和关键词字数

计算机的发展使计算单词出现的频次和索引(即在其他词汇的上下文中的单词数或上下文中的关键词[KWIC])都更容易了(Weber, 1990,pp.44-53)。这类分析(在语言学家中也被称为语料分析)可以帮助(采用诸如 WordSmith 这类语料分析软件的)文本数据分析过程迈出有价值的第一步。

## 主题内容分析

主题内容分析是内容分析中最常用的方法。在主题分析中,编码方案依据的是为把握文本主题而设计的各种类别。遗憾的是,不可能有一种通用的编码方案:不同的文本强调的东西不同,在同一个文本中不同的研究者要寻找的内容也不同。这就是为什么进行完善的主题分析需要十分熟悉输入的文本和它的特点。它还要求对编码方案进行大量的事前测试。

## 指涉内容分析

主题内容分析是梳理文本主要脉络的好工具。但意义也是其他各种语言游戏(language games)的结果:背景和前景信息、沉默和强调,或者描述同样事物的不同方式。就捕捉语言意义生成的复杂性而言,指涉内容分析是比主题分析更为合适的工具。克里彭多夫(Krippendorf,1980)告诉我们,指涉分析用于“探知一个现有的现象是如何被描写的任务”。这种描写受名词、形容词甚至不同的句法结构(如被动和主动语态的选择)的影响(Franzosi,2003)。

## “故事文法”和叙事结构

从这个技术发展最早期开始,研究者就希望找到一个以文本的语言特点为基础的内容分析方法(如 de Sola Pool, Hays,

Markoff, Shapiro, and Weitman; 参见 Franzosi,2003,Chap.1)。近年来,一些作者试图打通语言学和内容分析,提出了分析叙事文本的“故事文法”(又称文本文法或语义文法)这个基本的变体(Abell,1987; Franzosi,2003; Shapiro,& Markoff,1998)。从根本上说,故事文法无非就是简单的新闻五要素(5W)结构:何人(who)、何事(what)、何地(where)、何时(when)和何故(why)(以及如何);某人在时间和空间做某事,支持或反对别的某人;或主—谓—宾,或者更好一点,施事者—行动—受事者/受益者及其修饰语(例如,主语与宾语的类型和数量、时间和空间、行动的原因和结果)。这个技术在对社会运动和抗议运动的研究中发挥着重要作用。

## 内容分析如何进行

尽管量化也许不是内容分析正式定义必要的组成部分,但各种不同的内容分析方法的共同点还是与数字有关——这一点使它们区别于分析文本和符号材料的典型的定性方法,例如,叙事分析【Narrative Analysis】、谈话分析【Conversation Aanalysis】、话语分析【Discourse Analysis】和符号学【Semiotics】。但是,如果问题果真这样,那么内容分析又是如何把文字转换成数字的?转换的引擎肯定是“编码方案”,虽然在这个过程中起关键作用的是其他的“科学”程序:抽样【Sampling】、效度【Validity】和信度【Reliability】。

## 编码方案

编码方案是一套用于某个文本集的所有编码类别,其中每个编码类别都标识了一个研究者感兴趣的特点。编码方案被系统地用于所有被选中的文本,以便提取出统一的、标准化的数据:如果一段文本包含编码方案所含的任何编码类别的信息,相关的编码类别就被编码员“用记号勾出”(这个过程在内容分析中被称为编码【Coding】;现在的定性分析软件如 NUD * IST/N6 或 ATLAS.ti 允许使用者直接在计算机上对文本进行编码,并在编码进行时生成编码类别)。当编码完成后,每个编码类别得到的勾就被加总起来。最后,内容分析从文本变成数字的神秘魔法很简单:你只要计一下数就可以了。无论是文字、主题、指涉,还是行动者及其行动,这取决于具体的内容分析的类型,但无论何种类型,其数字都是计数得来的结果。

## 抽样

抽样给我们提供了一种得到研究结果的高效的好方法。研究者可能不会去研究所有可能得到的资源和文献,而会抽取恰当数目的资料来源和恰当数目的文件。抽样可能特别受内容分析的欢迎,因为内容分析需要耗费大量劳动,因此也是很昂贵的技术[而且抽样可以通过接受的抽样方案(acceptance sampling schemes)来检验编码数据的可靠性](Franzosi,2003)。通过抽样,你可以专注于报纸、信件或者其他感兴趣的子集合;你可以只关注被选的年份。甚至就被选年份而言,你也可以选读一份报纸一星期中的一两天,一个月中的一两周,或是一年中的一两个月的内容。建立一个不会让数据有偏的抽样框需要对输入材料有全面的了解。例如,只依靠报纸周日版的议题会对数据有什么影响?为了回答这个问题,以及与此类似的问题,任何数据来源和抽样框的选择都应在系统地比较分析的基础之上再进行。你的数据的效度(即你试图测量的概念和你实际的测量之间的基本一致性)是你采用的数据来源和抽样框的函数。

## 编码员信度

在内容分析中，信度问题（即测量的可重复性）被称作编码员信度。不同的编码员遇到了同样的文本，在使用同样的编码方案的情况下，是否会“勾选出”同样的编码类别？对那个问题的回答取决于编码方案和编码类别的设计：类别越抽象、定义越是理论化，不同的编码员就越可能得到不同的编码结果。让不同的编码员对同一材料进行编码，是检验每个编码类别信度的好办法。

## 一个例子

内容分析可以用于各种文本（例如，报纸社论、演讲、文件、信件、民族志田野笔记、深度访谈或焦点小组访谈的录音整理稿）和影像（例如，广告、照片）。下面是一组焦点小组访谈的录音整理稿，参加这组焦点小组访谈的都是出生在英国的巴基斯坦妇女，而她们的丈夫都来自巴基斯坦。这些参与者分别是：Nadia（N）、Laila（L）、Ayesha（A）以及主持人（M：Maria Zubair，我很感谢她整理了录音）。

M：你们的丈夫在这儿过得怎么样——当他们刚到这儿的时候？

L：我想这真的得看他们怎么生活……当他们在巴基斯坦的时候。我的意思是有些人，你知道的——他们没有其他人那么幸运，他们住在村子里或是其他什么地方，你知道的，他们其实没上大学，但也有一些人住在城里，有良好的教养——你知道，所以当他们来到这里的时候他们不会——他们不会明显地——他们相当不同但也不是那么不同……

L：对，我是说我的丈夫……当他来这里时，他是独子——嗯——他上过大学，高校，什么都上过，所以他有教养——所以他来这里的时候他没觉得有太大的不同……所以当然他们觉得不同，因为他……很不同，不是吗？

L：你知道的，到外面去——他们发现很难交朋友，我觉得这是主要的问题。

A：对。

L：交朋友——或者找个可去的地方——你知道的，有时候晚上或是周末我们说，哦，我们该去哪儿，但你知道，没地方可去——俱乐部和酒吧都是给英国人开的，不是给我们开的。

L：总之，他们觉得困难，因为他们是来这里，还要混迹于——嗯，英国出生的亚洲人中间——很难——更难的是，因为他们的思考方式不一样，你知道的，我是说我的丈夫——他以巴基斯坦的方式思考，而这里的男孩思考方式不一样——可不是吗！

A：我发现——我发现我丈夫也是一样。

L：对。

A：有时候我就是坐在那儿，嗯——我们没有（笑）——我们的思考方式不一样……他想事情都是巴基斯坦的方式……他回家的时候，他就坐在它（电视）前面——在沙发上——就在那儿。然后他就不动了。他就坐在那儿，就是这样，你知道的。（笑）

L：我丈夫曾经对我说，他来这儿的时候“哦，我不喜欢”……“哦，我不喜欢你穿英国的衣服什么的——哦，为什么你不戴你的头巾”，你知道，都是这些话。“为什么你不能穿亚洲的衣服去工作”……现在他从不那样说，你知道（N：对）。所以他已经适应了……但当他刚来这里时他们是受了点文化震荡——我的意思是什么都是不同的（M：对）和那里比起来的话，你知道的，所以——那里是不同的，对吗？完全不同。

A：（笑）我的丈夫——他，你知道的，他

就是让我紧张。当他才到这儿时他曾经穿着灯笼裤[巴基斯坦服装]——我跟他说,你出去看看……你穿这个出去不行。

甚至单词的计数也可以揭示出社会背景的二分性(考虑版主的引导性问题,也许这是在意料之中的)。出现得最频繁的词是他们(21)、他(16)、我(12),还有那里(6)。像不同(5)和难(5)这样出现频率比较高的词在突出显示英国出生的巴基斯坦妇女和巴基斯坦男子之间婚姻的复杂社会关系时,也还有很长的路要走。毕竟像相同、简单这些词都从没有在文本中出现过。主题编码类别也可以有助于梳理这些不同世界的特性。无论是借助专业的计算机软件,还是只借助纸张,以主题类别为基础的编码方案都必须处理社会生活的以下几个方面:家庭、友谊、工作、休闲。但是这些类别太宽泛了。这个编码方案必须聚焦于“婚姻关系”中的“难/易”“我们的家庭和岳父岳母(公公婆婆)”“男性朋友和女性朋友”,以及“在巴基斯坦的朋友和在英国的朋友”。这些类别的频数分布可以让我们更好地理解“你们的丈夫在这儿过得怎么样”这个问题。

但文本的含义还有很多。先来看 Laila 的话“有些人……他们没有其他人那么幸运”。对词汇的选择不是随意的。想一想替换成“没那么勇敢”“没那么发奋”“没那么爱冒险”。不对——他们就是“没那么幸运”。对他们没有指责,就是碰运气。保守的精英世界观令我们相信,我们手中的(扑克)牌都是一样的。我们用它们来为自己做什么,则取决于我们自己,勇敢、大胆和创业精神。而 Laila 组织话语的方式却含蓄地遵循着更为自由主义而非保守主义的脉络。但她也以隐含的社会科学模型为基础来组织她的话语,我们可以轻松地识别出来。在村子里长大的人不上大学。在城里的人教养良好。他们受教育。他们没那么不同。社会科学的城市化和教育与耐受性关系模型隐含其后。最后,Laila 用一个非常二元的词汇“我们”和“他们”来组织其话语:“俱乐部和酒吧都是给英国人开的,不是给我们开的”,“他以巴基斯坦的方式思考”,“我们的思考方式不一样”。Laila、焦点小组里的其他妇女以及她们的丈夫和英国人不同。但是她们也和自己的丈夫不同。而 Laila 受过教育的丈夫与其他丈夫中的有些人也不同。

这些不同的论证和组织模式可以通过如下的编码类别凸显出来:“保守的政治观点”以及“自由的政治观点”,“包容性的世界观”和“排斥性/二元的世界观”,还有“推理(或论证)模式”,以及诸如“情绪性的”“社会科学的”等子类。对修辞和论证模式的强调会让我们加入诸如“诉诸宗教”“诉诸科学”“诉诸情感”“诉诸理性”“诉诸平等的价值观”等类别。

指涉分析可以被用于处理对社会行动者和社会情境的不同描写。对每个主要的行动者(例如,英国人、丈夫们、妻子们、朋友们、在巴基斯坦的亲戚们)、事件(例如,回巴基斯坦、结婚、搬到英国)或情境(在家的一个晚上、周末的时光、穿衣服和吃饭的习惯),我们可以有三种基本的指涉编码类别,即积极的、消极的和中性的。

## 结论

内容分析作为一种研究文本的典型的定量化手段,为梳理文本或其他任何符号材料的意义提供了一种宝贵的工具。如果遇到要分析上述种类的几百页的录音整理稿(或者其他类的文件/符号材料),各种内容分析技术肯定会有助于我们揭示出资料中的模式。定量分析的一个优势在于它很有效:它有助于我们处理大量数据,使我们不会被庞大的数据淹没。它的缺点在于可能会遗漏意义生产过程中的细微差别。即使上文已经讨论过的为数不多的几个例

子，也应使问题变得很清楚。考虑用定量化方法还是定性化方法来处理文本，这样的思维方法是一种误导。每种方法对文本分析的理论/方法论的发展都有它的长处和短处（想想这个：对几百页的录音整理稿作定性化分析要花多长时间？你如何保证你没有依赖最好听的那些话，或者你什么也没有忘记？），加上计算机软件技术的发展，正在加速模糊定量和定性方法的界限［例如，计算机辅助定性数据分析软件（CAQDAS）］（Fielding & Lee，1998；Kelle，1995）。就当前技术发展状况来说，最好的方法可能是将绘制宏大的社会现象图景的定量内容分析法和对精选的文本进行鞭辟入里的探究的定性内容法融为一体的组合分析法。

——Roberto Franzosi

（杨可译　沈崇麟校）

## 参考文献


Abell, P. (1987). *The syntax of social life: The theory and method of comparative narratives.* Oxford, UK: Clarendon.

Berelson, B. (1954). Content analysis. In G. Lindzey (Ed.), *Handbook of social psychology* (Vol. 1, pp. 488-522).Reading, MA: Addison-Wesley.

Fielding, N., & Lee, R. (1998). *Computer-analysis and qualitative research.* London: Sage.

Franzosi, R. (2003). *From words to numbers.* Cambridge, UK:Cambridge University Press.

Holsti, O. R. (1969). *Content analysis for the social sciences and humanities.* Reading, MA: Addison-Wesley.

Kaplan, A. (1943). Content analysis and the theory of signs.*Philosophy of Science*, *10*, 230-247.

Kelle, U. (Ed.). (1995). *Computer-aided qualitative data analysis.* London: Sage.

Krippendorf, K. (1980). *Content analysis: An introduction to its methodology.* Beverly Hills, CA: Sage.

Shapiro, G., & Markoff, J. (1998). *Revolutionary demands:A content analysis of the Cahier de Doléances of 1789.* Stanford, CA: Stanford University Press.

Weber, R. P. (1990). *Basic content analysis.* Newbury Park,CA: Sage.


# 前后关系效应（Context Effects）

参见次序效应【Order Effects】。

# 情境效应（Contextual Effects）

在社会科学中团体这一概念起着重要的作用。个人拥有某个团体的成员资格是否会对其思想和行动有什么特别的影响？居住在一个由许多民主党人构成的街坊是否会使得个人比那些不住在那样的街坊的个人有着更高成为一个民主党人的概率？诸如这样一种由团体造成的影响称为情境效应。对这些效应所做的研究称为情境分析或多层次分析【Multilevel Analysis】。有时它也被称为层级【Hierarchical】模型研究。

如果情境效应存在，那么如下这样的问题就会发生：为什么属于一个团体便会有影响？这种机制是如何运作的？我们究竟应如何来测量这些影响？前两个问题最好留

给心理学家来解答,而第三个问题则会引起统计学家的极大兴趣。特别是教育问题中,一个教室将会构成一个显而易见的情境分析的团体。

当有来自几个组的个体有两个变量的数据时,在每个组内可以分析因变量 $Y$ 和自变量 $X$ 的关系。回归分析给出了每个组的回归线的斜率【Slope】和截距【Intercept】。这样就可能会出现组与组之间的斜率和截距是变化的这种情况。如果回归线的这种变化大于随机性独自可以解释的变化,那么必定有什么与组本身有关的东西在影响着回归线。

导致回归线发生变化的原因可能有若干种。一种可能是二者都取决于通过组均值的组的 $X$ 变量水平。因此,一种可能的情境模型可表达为如下的方程:

$$a_j = c_0 + c_2\bar{x}_j + u_j$$
$$b_j = c_1 + c_3\bar{x}_j + v_j$$

式中的 $a$ 和 $b$ 是第 $j$ 组的截距和斜率,而 $u$ 和 $v$ 则表示残差项。也可用第 $j$ 组的直线来替代方程中的 $a$ 和 $b$。这样便可得到如下的方程:

$$y_{ij} = c_0 + c_1x_{ij} + c_2\bar{x}_j + c_3(x_{ij}\bar{x}_j) + e_{ij}$$

它是第 $j$ 组中的第 $i$ 个个体。因此,现在可以对来自所有的组的数据进行多元回归【Multiple Regression】,并得到 $c$ 系数。对于这个模型而言,$c_1$ 测量的是个体水平上的 $X$ 的效应,$c_2$ 测量的是组水平上的 $X$ 的效应,而 $c_3$ 测量的则是个体和组水平效应的交互项在 $Y$ 上的效应。使用这种分析存在一个问题,那就是残差【Residual】$E$ 取决于 $X$ 的组均值;此外,还有另外一个问题,那就是存在于三个解释变量中的共线性【Collinearity】问题。

情境分析不限于组的截距和斜率取决于 $X$ 的均值这样一种情形。它们也可能取决于各种各样其他的变量。

贝叶斯统计法尤其适用于多水平分析,因为各个组的斜率和截距是变化的。这样我们就有可能详细说明服从特定分布的组内截距和斜率。

——Gudmund R. Iversen

(沈崇麟译校)

## 参考文献


Bryk, A.S., & Raudenbush, S.W. (1992). *Hierarchical linear models*. Newbury Park, CA: Sage.

Goldstein, H. (1987). *Multilevel models in educational and social research*. London: Griffin.

Iversen, G.R. (1991). *Contextual analysis* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-081). Newbury Park, CA: Sage.


## 列联系数(Contingency Coefficicent)

参见对称量度【Symmetric Measures】。

## 列联表(Contingency Table)

列联表是一种根据两个或更多个变量的分类将观察数据的频数分类的统计表。

由两个变量交叉分类形成的表称为双向列联表(two-way contingency table)(实例请参见表 1),三个变量的表称为三向列联表(three-way contingency),而一般来讲,三个或三个以上的表都称为多向列联表(multiway contingency table)。多向列联表的分析有时也称为多变量列联表分析(multivariate contingency analysis)。

在一张表的所有格子中,有的格子可能没有观察的计数。有时,我们把列联表中含有零计数的格子称为空格。零计数可能由两种不同的机制产生。如果零计数是由样本容量引起的(如伊利诺伊州的美籍华裔农民),那么这样的零就称为抽样零。如果,一个格子在理论上就不可能有观察(如男性子宫帽用户),那么这样的零称为结构零。含有至少一个结构零的列联表称为不完整表。在列联表中有许多格有较低的频数时,不论它们是零或者不是零,诸如这样的表都称为稀疏表【Sparse Table】。

## 历史发展

20 世纪初,早期的定类数据分析【Categorical Data Analysis】主要都采用列联表分析。在卡尔・皮尔逊和乌迪・耶尔之间的著名争论引发了人们对于列联表分析的兴趣。我们应该怎样来分析一张 2 × 2 的列联表呢?皮尔逊坚定地相信,连续的双变量分布构成了交叉分类表的观察计数。耶尔则认为,含有诸如接种与不接种这样的离散类别的变量,最好作无假定的潜在分布的离散变量处理。皮尔逊的列联系数是依据坍缩为离散交叉分类的,近似的潜在双变量正态分布的相关来计算的。而耶尔的 $Q$ 则是直接用观察的离散数据的优势比【Odds Ratio】函数计算的。

整个 20 世纪在这方面取得的成就常常可以与那些发明列联表分析的统计量和检验的名字联系起来。Pearson 卡方检验, Yule $Q$, Fisher 精确检验, Kendall tau($\tau$), Cramér $V$, Goodman 和 Kruskal tau($\tau$) 和 Cochran-Mantel-Haenszel 检验(或 Mantel-Haenszel 检验),这些只不过是其中的一些例子而已。今天的统计软件一般都包含这些统计量(例如,SAS 中的 proc freq 就有上面提到的大多数统计量)。

**表 1 父母社会经济地位(SES)和精神健康状态**
**(括弧中的数字是期望值)**

| 父母社会经济地位(SES) | 精神健康状态 | | | |
|---|---|---|---|---|
| | 好 | 轻度症状 | 中度症状 | 精神损伤 |
| A(高) | 64(48.5) | 94(95.0) | 58(57.1) | 46(61.4) |
| B | 57(45.3) | 94(88.8) | 54(53.4) | 40(57.4) |
| C | 57(53.1) | 105(104.1) | 65(62.6) | 60(67.3) |
| D | 72(71.0) | 141(139.3) | 77(83.7) | 94(90.0) |
| E | 36(49.0) | 97(96.1) | 54(57.8) | 78(62.1) |
| F(低) | 21(40.1) | 71(78.7) | 54(47.3) | 71(50.9) |

列联表分析可视为当代分类数据分析的基础。我们可以把20世纪晚期对数线性模型【Log-Linear Modeling】的一系列发展看成列联表分析的发展和细化。同样,潜质和潜类分析的进展也是与列联表分析的统计基础不可分的。

## 例子

我们给大家提供一张经典的、涉及中镇曼哈顿(midtown Manhattan)精神健康与父母社会经济地位(SES)的数据表,该表的数据曾经过许多人(如里奥·戈德曼和克洛格等人)的分析(见表1)。

为了分析列联表,我们需要有期望频数。令 $F_{ij}$ 为第 $i$ 行第 $j$ 列的观察频数 $f_{ij}$ 的期望值。独立模型(SES和精神健康没有关联)的 $F_{ij}$ 由

$$F_{ij}=\frac{f_{i+}f_{+j}}{f_{++}}$$

给出。该式中的 $f_{i+}$ 给出了第 $i$ 行的列总计数,$f_{+j}$ 给出了第 $j$ 列的行总计数,而 $f_{++}$ 则给出了整个表格的总计数。例如,$f_{1+}=64+94+58+46=262$,$f_{+1}=64+57+57+72+36+21=307$ 和 $F_{11}=(262\times307)/1\ 660=48.5$。我们计算了这个表格相应的期望值,并把它们列在了表1每一格的括弧中。

虚无假设是两个变量没有关系或相互独立的。为了检验这个假设,我们计算了Pearson卡方统计量和最大似然比统计量 $L^2$(或 $G^2$)的值:

$$\chi^2=\sum_i\sum_j\frac{(f_{ij}-F_{ij})^2}{F_{ij}}$$

和

$$L^2=2\sum_i\sum_j f_{ij}\log\left(\frac{f_{ij}}{F_{ij}}\right)$$

应用这些公式,我们求得Pearson卡方统计量的值为45.99,而 $L^2$ 的值为47.42。在有15个自由度(该值等于行数减1乘列数减1),在任何常规的显著水平下,我们拒绝独立性虚无假设,并得出精神健康和SES是相关联的结论。独立性模型也可以从用期望频数计算的概率比率角度来理解。为了理解独立性假定,我们可以计算列联表中每个相邻的 $2\times2$ 表的交叉乘积的比。那就是 $(48.5\times88.8)/(95.0\times45.3)\approx1.0$,$(95.0\times53.4)/(57.1\times88.8)\approx1.0$,…,$(57.8\times50.9)/(62.1\times47.3)\approx1.0$。我们很快就会发现所有这些优势比都近似于1!优势比【Odds Ratio】的值为1说明形成表格的两个变量之间是没有关系的。

两个变量本质上都是有顺序的,但是目前的分析却未曾考虑这一信息。为了分析两个变量间的这种类型的关联,则必须用对数线性模型对模型作出进一步调整。

## 列联表的坍缩性

列联表的突出问题是坍缩性问题。简单地说,如果一种特定的统计模型的拟合不受合并它的维度(即删除变量)、合并变量中的某些类别(即减少答案的类别数)或同时合并变量和类别二者的显著影响,那么这张列联表就是可坍缩的。在目前的这个例子中,有人可能怀疑在轻度和中度症状之间只有很小的差别,注意到表1的中间两行的观察值和期望值之间差别都同样低于1。表2列出了将表1中间两类坍缩之后的观察数据(和期望计数)。

表 2 父母社会经济地位(SES)和精神健康列联表

| 父母社会经济地位(SES) | 精神健康状态 | | |
|---|---|---|---|
| | 好 | 中度症状 | 精神损伤 |
| A(高) | 64(48.5) | 152(95.0) | 46(61.4) |
| B | 57(45.3) | 148(88.8) | 40(57.4) |
| C | 57(53.1) | 170(104.1) | 60(67.3) |
| D | 72(71.0) | 218(139.3) | 94(90.0) |
| E | 36(49.0) | 151(96.1) | 78(62.1) |
| F(低) | 21(40.1) | 125(78.7) | 71(50.9) |

从表 2 中可以得到一组新的检验统计值。基于六行、三列和 10 个自由度的皮尔逊卡方,43.51 和 $L^2$,44.94(在 0.05 水平,结果仍然是显著的)。取新值和老值之差,无论是基于皮尔逊卡方统计值,还是 $L^2$统计值的似然比检验都将给出有 5(15-10)个自由度的检验统计值 2.74。这个结果服从卡方分布,是高度不显著的,说明这中间两个中度症状类是可以坍缩的,不会因此而丢失太多的信息。读者可自行尝试一下父母社会经济地位(SES)类别是否可以坍缩。

——Tim Futing Liao
(沈崇麟译校)

## 参考文献


Agresti, A. (1990). *Categorical data analysis*. New York:John Wiley.

Fienberg, S. E. (1981). *The analysis of cross-classified categorical data* (2nd ed.) Cambridge:MIT Press.

Rudas, T. (1997). *Odds ratios in the analysis of contingency tables*. Thousand Oaks, CA:Sage.


# 他定效应(Contingent Effects)

在对自变量【Independent Variable】和因变量【Dependent Variable】的关系进行理论化时,有时可能会遇到这种关系根据另一个自变量的值而定的情况。这种类型的关系蕴含着一个设计用于经验上估计这些关系的,在任何一种模型中的一个乘法项。用正规的数学语言叙述,这种模型将取如下的形式:

$$Y = B_{12}(X_1 \cdot X_2)$$

式中的 $Y$ 是因变量,$X_1$ 是自变量,$X_2$ 是决定 $X_1$ 和 $Y$ 的关系的变量,而 $B_{12}$ 则是定义这种关系的参数。因此,$X_1$ 和 $Y$ 之间的值取决于 $X_2$ 的值。这方面的例子来自希布斯(Hibbs,1982)在越南战争期间进行的一个有关美国总统声望的研究。希布斯将个人对总统的工作的看法理论化,认为它将直接受到最近几周越南战场上美国军人伤亡数的影响。此外,他还提出一种理论,声称这种效应的量将取决于个体的社会阶层,因为各个阶层的伤亡人数是不成比例的,伤亡主要集中在那些较低的社会阶层。

他定效应的第二种定义来自标准的多变量模型推论。例如,下列是一个有两个自

变量的多变量普通最小二乘【Ordinary Least Squares】回归中的非随机部分：

$$Y = B_0 + B_1X_1 + B_2X_2$$

式中的 $Y$ 是因变量，$X_1$ 和 $X_2$ 是自变量，$B_0$ 是一个与其他感兴趣的参数一起估计的截距项。$B_1$ 是 $X_1$ 在 $Y$ 上的，依 $X_2$ 在 $Y$ 上的效应而定的效应（要估计的）参数。$B_2$ 是 $X_2$ 在 $Y$ 上的，依 $X_1$ 在 $Y$ 上的效应而定的效应（要估计的）参数。还有另一种常用的表述法，那就是每个斜率参数（在本例的模型中，$B_1$ 和 $B_2$ 就是斜率参数），就是"控制了"模型所有其他的变量时的一个自变量在因变量上的效应。

——Guy Whitten

（沈崇麟译校）

＊也可参见交互效应【Interaction Effect】、乘积【Multiplicative】。

## 参考文献

Hibbs, D.A. (1982). The dynamics of political support for American presidents among occupational and partisan groups. *American Journal of Political Science*, 26, 312-332.

# 连续变量（Continuous Variable）

连续变量是这样一种变量——在其值域内，任何值都是可能的。有时，它也称为测度变量（metric variable）或定量变量（quantitative variable）。在实际研究中，整数变量也常被视为是连续的。

——Tim Futing Liao

（沈崇麟译校）

＊也可参见测量层次【Level of Measurement】。

# 对比编码（Contrast Coding）

对比编码是一种为 $j-1$ 个虚拟变量【Dummy Variable】编码，以体现一个有 $j$ 个类别的定类变量【Nominal Variable】含有的信息。对比编码的虚拟变量使研究者得以为比较的目的而将类别聚合成有意义的分组的同时，检验确定每个组的那些组成类别是否彼此显著不同。对比编码虚拟变量系数给我们提供了顺序分解自理论上感兴趣的组合的子组的均值比较。

为了说明这一问题，我们来考虑一下婚姻状况（已婚、分居、离婚、鳏寡和从未结过婚）和小时工资（以美元计）之间的关系。在这个例子中，婚姻状况由五个（$J$）子组组成，由四个对比类（$J-1$）代表。虽然有若干个初始聚合都是合理的，但我们将第一个对比构建为一个目前那些在法律上已婚（已婚和分居）的人和那些目前为未婚的（从未结过婚、离婚和鳏寡）的人之间的对比，具体则如表 1 所列。每个相继的对比，将把这两个组分解并组成子组。对比 2 将比较分居的回答人和已婚的回答人。因为"未婚聚合"含有三个子组，因此选择在对比 3 中比较"从没有结过婚"的回答人与离婚和鳏寡的回答人。在对比 4 则比较离婚的回答人和鳏寡的回答人。

在将对比编码用于虚拟变量时，还有另外两条规则适用。第一条，任何单独一个对比编码变量内的编码之和必须为零。第二

条,跨虚拟变量的编码必须正交,正如表 1 的底部冠以“独立性”的那一部分所示。

表 1

| 兴趣组 | 新变量状态 | | | | |
|---|---|---|---|---|---|
| | 对比 1(C1) | 对比 2(C2) | 对比 3(C3) | 对比 4(C4) | 均值(以美元计) |
| 已婚 | −1/2 | −1 | 0 | 0 | 15.33 |
| 从未结过婚 | 1/3 | 0 | −1 | 0 | 12.95 |
| 分居 | −1/2 | 1 | 0 | 0 | 11.64 |
| 离婚 | 1/3 | 0 | 1/2 | 1 | 12.94 |
| 鳏寡 | 1/3 | 0 | 1/2 | −1 | 11.93 |
| 总和 | −1/2+1/3−1/2 +1/3+1/3=0 | −1+0+1 +0+0=0 | 0−1+0+1/2 +1/2=0 | 0+0+0 +1−1=0 | 总均值 14.26 |
| 独立性 | C1 和 C2=(−1/2)(−1)+1/3(0)+(−1/2)(1)+(1/3)(0)+(1/3)(0)=0<br>C1 和 C3=(−1/2)(0)+(1/3)(−1)+(−1/2)(0)+(1/3)(1/2)+(1/3)(1/2)=0<br>C1 和 C4=(−1/2)(0)+(1/3)(0)+(−1/2)(0)+(1/3)(1)+(1/3)(−1)=0<br>C2 和 C3=(−1)(0)+(0)(−1)+(1)(0)+(0)(1/2)+(0)(1/2)=0<br>C2 和 C4=(−1)(0)+(0)(0)+(1)(0)+(0)(1)+(0)(−1)=0<br>C3 和 C4=(0)(0)+(−1)(0)+(0)(0)+(1/2)(1)+(1/2)(−1)=0 | | | | |

一般来说,设计这些虚拟变量的描述性统计量和零阶相关都没有什么用(Hardy,1993)。在对比编码的虚拟变量用作普通最小二乘回归【Ordinary Least Squares Regression】的自变量时,常数给出的是子组均值的未加权均值,计算方法是 $[\sum_{j=1}^{J}\overline{X}_j]\div J$,与用效应编码的虚拟变量的回归一样,但却与二进制【Binary】编码的虚拟变量的回归不同。回归系数【Regression Coefficients】则可用于计算对比内的均值差。计算使用的公式是 $C_j=B[(n_{g1}+n_{g2})/(n_{g1})(n_{g2})]$,式中的 $n_{g1}$ 是包含在第一个组内的子组数;$n_{g2}$ 是包含在第二个组内的子组数;而 $B$ 则是感兴趣的虚拟变量的回归系数。

小时工资对 4 个对比编码的虚拟变量,加上一个常数的回归结果列于表 2 中。常数项的值为 12.96,这个值可用子组均值的未加权均值验证。计算的方法是用表 1 提供的类均值之和除以 5。为了评估回归系数,我们来看一下对比 1 的系数。将它代入上面的方程,便会有下式:C1 = −1.054×((2+3)/(2 ×3))= −0.878。这个值是未加权的合法婚姻的子组未加权均值([15.33+11.64]/2=13.49),无合法婚姻子组的未加权均值([12.95+12.94+11.93]/3=12.61)之差。这就是说,无合法婚姻的人的小时工资平均比有合法婚姻的人平均少0.88个美元。然而这个系数的 $t$ 检验在常规水平上却是在统计上不显著的,因此,0.88 个美元的差

别是不可靠的。确定小时工资是否存在婚姻状态上的差别的恰当推论检验是 $F$ 检验，而不是 $t$ 检验，因为检验要求的有关婚姻状况的信息包括所有四个虚拟变量。

表 2

| 模型 | 未标准化系数 | | $t$ | 显著度 |
|---|---|---|---|---|
| | $B$ | 标准误差 | | |
| 常数 | 12.96 | 0.471 | 27.527 | 0.000 |
| 对比 1 | -1.054 | 0.993 | -1.061 | 0.289 |
| 对比 2 | -1.844 | 0.353 | -5.226 | 0.000 |
| 对比 3 | -0.346 | 0.772 | -0.448 | 0.654 |
| 对比 4 | 0.502 | 1.111 | 0.452 | 0.651 |

——Melissa A. Hardy
Chardie L. Baird
（沈崇麟译校）

＊也可参见虚拟变量【Dmmy Variable】、效应编码【Effects Coding】。

## 参考文献

Hardy, M.A. (1993). *Regression with dummy variables.* Newbury Park, CA: Sage.

# 控制(Control)

控制涉及一系列功能，借助这些功能的两个主变量，即自（解释）变量和因（或应）变量之间的关系可以在不受其他因素干扰的情况下得以确定。控制有四个主要的功能，分别是：固定或使混杂【Confounding】效应"保持不变"；检测变量之间的虚假效应；在实验检测中作为一个参照点和在其他变量也有效应时确定自变量的净效应。

混杂效应可能是那些双变量关系中的干扰效应，或者那些在这种关系前面的干扰效应。或者这种关系可能是一种虚假的关系，也就是说，有第三个变量同时在对自变量和因变量起作用。为了探测这样一种混杂因素的效应，我们对它们进行控制，并对同类项进行比较（参见其他条件不变【Ceteris Paribus】）。例如，我们也许会注意英国在校学生中的某个少数民族族群的学生的学习成绩比白人学生差。我们会因此而设想，是和种族关联的文化和遗传因素导致少数民族族群的学生的学习成绩比其他族群的学生差。然而，我们考虑到收入和贫困也同样会对在校学生的学习成绩有所影响，而某些少数民族族群都集中在收入档次中的低档，因此我们希望对收入

水平加以控制。在对收入水平加以控制之后(即将穷孩子与穷孩子比较,富孩子与富孩子比较),我们发现这一关系并非持续不变的,那些孟加拉裔的孩子的学习成绩至少不比同一收入水平家庭的白人孩子差。在控制了学生的家庭背景之后,种族与学习成绩差的表面关系便不再是持续不变的了。图 1 显示了一个干扰效应的实例。今天,孟加拉裔学生的成绩仍然较差,但人们对造成这一现象的推理过程已有所修正,已不再把它归结于学生的种族,而是归结于他们的家庭贫困。

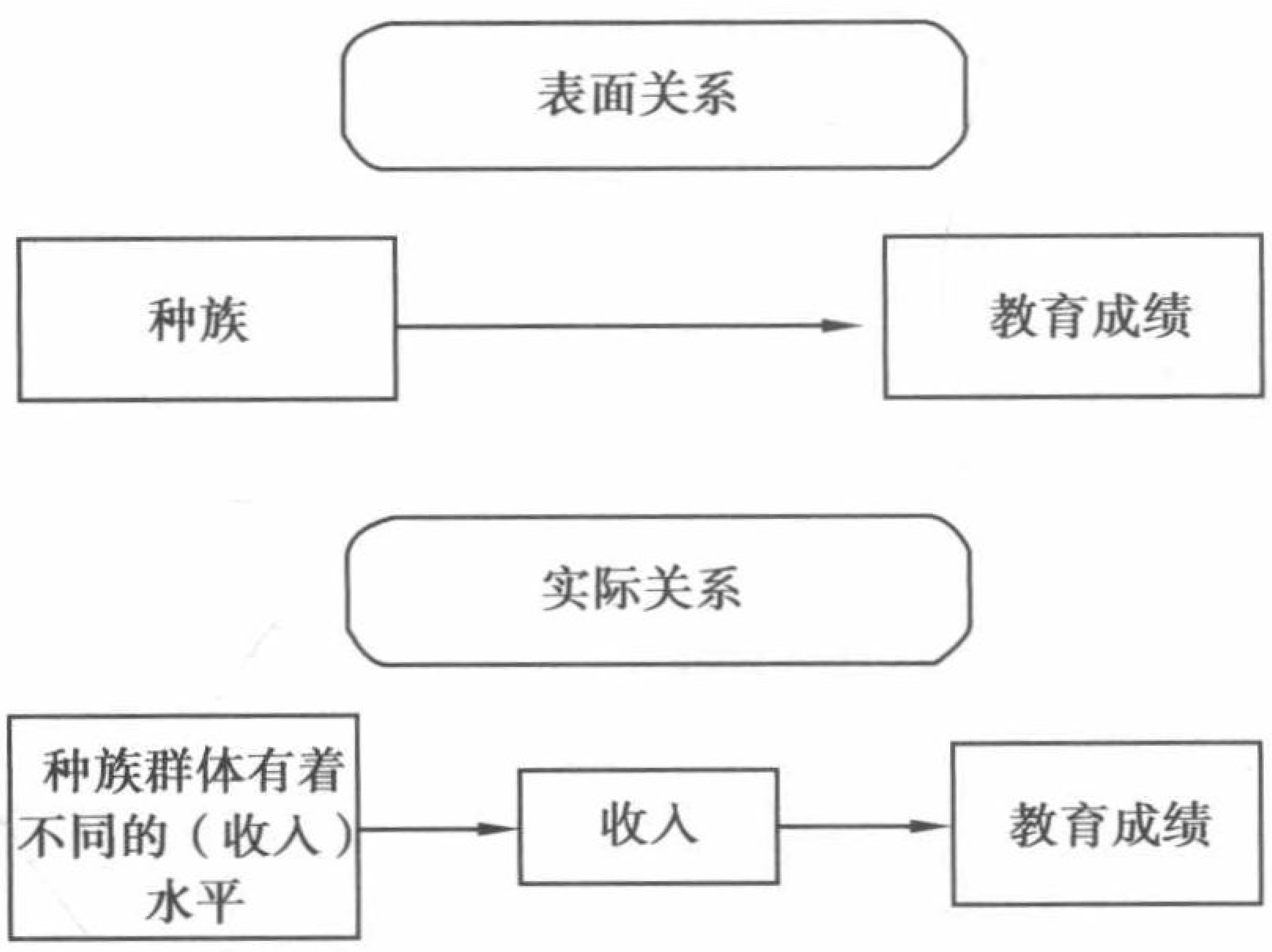



**图 1　在研究种族和教育成绩之间的关系时控制收入的效应**

我们用女性的平均工资低于男性作为一个例子来阐述干预关系【Intervening Relationship】问题。我们最初可以把它设想为性别和低工资之间的关系,这种设想注重工作场所内的性别歧视。与之相关的观点认为,在职业的层级结构中,女性趋于集中在它的较低层。如果使跨男性和女性的职业层次保持不变,是否仍然可以看到某种性别效应呢?我们发现,性别效应仍然存在,但是却要小得多。妇女的职业结构解释了相当一部分她们的低收入,而非她们的性别本身。而这一问题正是通过比较同一职业层次男性和女性的工资才被揭示的。图 2 为我们阐明了这样一种关系。

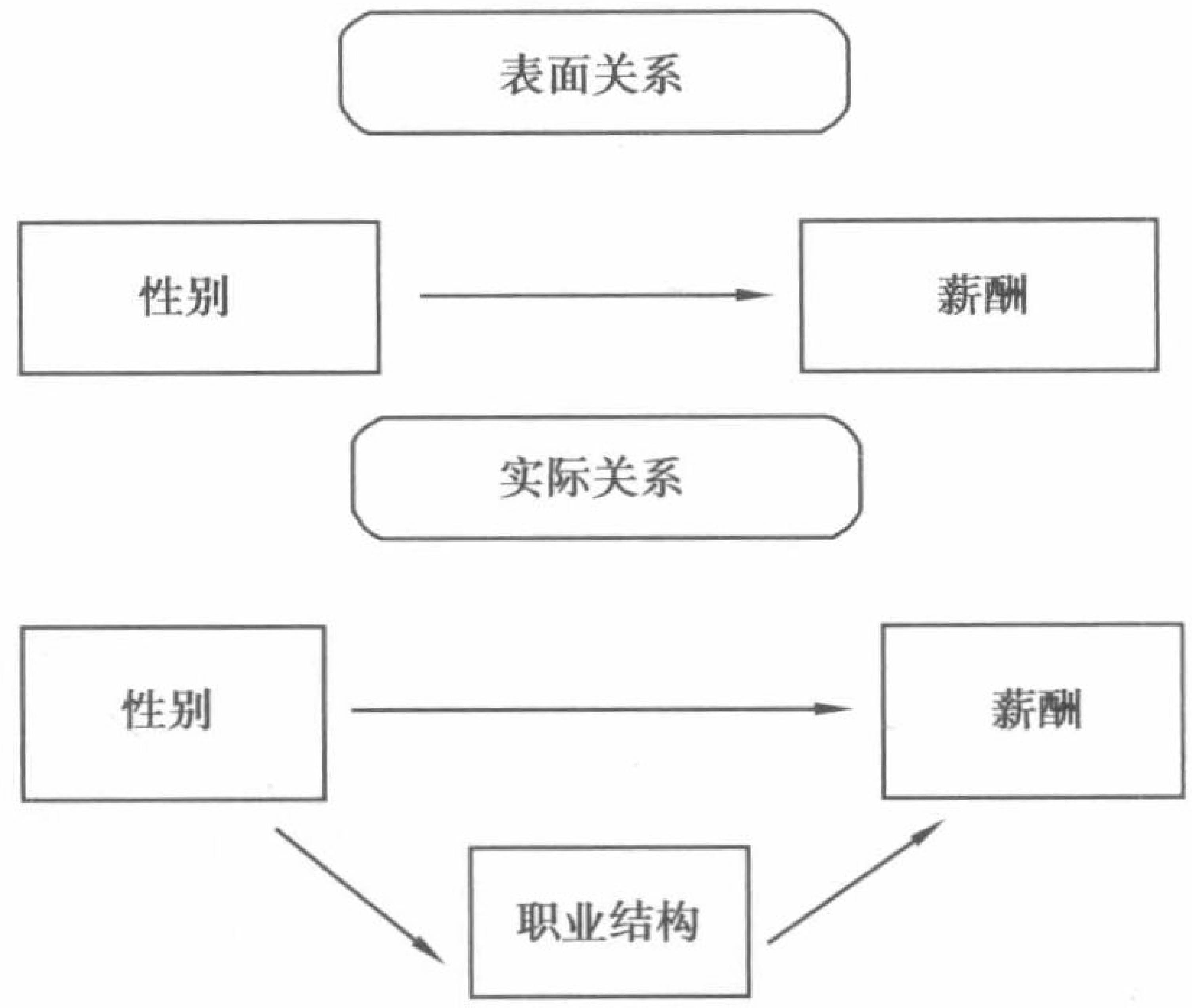



**图 2　在评估性别和收入的关系时,保持职业层次不变的效应**

另外一些相关的感兴趣的因素,可通过实验设计或用统计方法来控制。在实验设计中,干预效应可通过将干预随机地应用于测试的个案的某一个子组和对其余在控制组【Control Group】中的个案则不做干预来测试。对控制组也可进行挑选,以使之在关键方面(如年龄、性别等)与"实验组"匹配。例如,医疗干预一般都是用随机对照试验(randomized controlled trials)测试的。所谓随机是指那些实验的个案都是从感兴趣的总体中随机地选取的;而控制则是指每一个选来做实验【Experiment】的个案,都是用来作比较的,并不接受干预,但在其他每一方面的处理都与实验组一样,包括可能用安慰剂对其进行"治疗"。如果不仅实验组和控制组不知道自己究竟属于哪一组,而且实验的实际操作者本身也不知道施予的是干预(还是安慰剂),那么这样的实验就是双盲的,故而能在最大限度上缩小研究者偏倚。在这样的实验中,如果干预是随机分配的,那么在实验组和控制组的结果的任何差异

都可被归结于干预。

尽管实验方法在心理学中经常使用,但要把它在社会科学研究中普及似乎不仅不太实际,而且意义也不大。例如,建立一个父母分居对孩子学业影响的实验是不可行的。因为它要求有一个与其配偶分居的实验组的随机样本和一个属于他们的配对的控制组。相反,使用统计方法则可对混杂效应加以控制。我们有各种可同时评估许多不同变量的效应的多变量方法,因而在另一个(另几个)变量被控制时,我们能对净效应加以识别。回归分析,如多元回归分析【Multiple Regression Analysis】或 logistic 回归【Logistic Regression】被最为普遍地用于评估这样的净效应。在这样的分析中,其他自身并未显示任何显著效应的变量也被包括在了模型中,它们之所以被包括在模型中,只是为了确保所有相关的因素都已被控制而已。

——Lucinda Platt

(沈崇麟译校)

*也可参见条件比较法【Ceteris Paribus】、Logistic 回归【Logistic Regression】、多变量分析【Multivariate Analysis】、多元回归分析【Multiple Regression Analysis】。

## 参考文献

Agresti, A., & Finlay, B. (1997). *Statistical methods for the social sciences* (3rd ed.). Upper Saddle River, NJ: Prentice Hall.

## 控制组(Control Group)

控制组是一种和实验组(或治疗组)进行对比的组。它由一种没有接受治疗,但其效应也在研究考虑之列的研究单位或一种确实服用了安慰剂,但效用不在研究考虑之列的研究单位组成。控制组有时也称为基线组或对照组。它们或通过像组间设计(between-subject designs)那样随机地分配研究单位,或像组内设计【Within Subject Designs】那样由研究单位自己形成的。在一个真正的实验研究和组间设计中,研究单位(如儿童、诊所的病人、学校、医院、社区等)是先从各自的总体进行随机选择的,然后再将选出的单位,随机地分配给控制组或实验组。在研究结论中,结果量度(一个或多个因变量上的得分)将在控制组和实验组之间进行比较。治疗的效应将根据控制组和一个或多个实验组之间的差(或若干个差)来评估。虽然在真实的实验研究和组内设计中,研究单位也是随机地从各自的总体中选取的,但却是非随机地分配给控制组与实验组。相反基线数据是从研究单位自身收集的。这些数据被作为"控制数据"处理,用来与假设为治疗结果的结果量度作比较。因此,在组内设计中,研究单位自己充当了自己的"控制组"。对准实验【Quasi-Experiment】研究而言,治疗并非像真正的实验研究那样,由实验者亲自施予研究单位或参与者,治疗(或者称为处理)则被广泛地理解为参与者具有的特征,如性别、年龄和社会经济地位(SES),或他们所处环境的性质,如公立与私立学校及单亲和完整的家庭背景等。在研究者几乎无法控制把参与者分配给各种不同的治疗条件时,准实验研究则也可能是必需的。在这种情况下,嵌套设计则是恰当和必需的。

——Chao-Ying Joanne Peng

(沈崇麟译校)

*也可参见实验【Experiment】。

## 参考文献

Huck, S. (2000). *Reading statistics and research* (3rd ed.). New York: Addison-Wesley Longman.

Kirk, R.E. (1995). *Experimental design: Procedures for the behavioral sciences* (3rd ed.). Belmont, CA: Brooks/Cole.

Maxwell, S. E., & Delaney, H. D. (1990). *Designing experiments and analyzing data: A model comparison perspective*. Belmont, CA: Wadsworth.

# 方便样本(Convenience Sample)

诚如它的名称所言,方便样本采用的样本单位都是研究者马上就可以得到的。有时它也称为随意抽样(accidental sampling),因而它是一种非概率抽样【Nonprobability Sampling】方法。这就是说,总体中的每一个成员有着未知和不等的被选取的概率。

方便样本的优点在于费用比较低廉,按照它的定义,样本也比较容易得到。有时这样一种抽样形式是一种最为有效地进入那种难以进入的总体,如性工作者或非法药物的使用者总体的途径。在感兴趣的总体【Population】难以定义,因而无法使用任何可靠的随机抽样形式时,方便抽样便成一种比较恰当和可以接受的抽样方法。在难以取得进入感兴趣总体的通道时,它也会被看成是一种合理正当的抽样方法。

尽管使用方便样本的研究也能取得令人感兴趣的调查结果,但是它们的不足之处在于无法进行超样本的概括。研究者不知道样本是否代表了研究的总体。只依靠一个方便样本,研究者将无从得知它们的样本对其他的群体究竟是典型的还是不典型的。然而,这种缺点是所有非概率样本所共有的。研究者在使用这种形式抽样时,还应考虑它涉及的潜在偏倚【Bias】。

方便样本的使用不限于任何特殊的主题范围。研究者曾用方便样本来处理老年歧视问题,考察市中心黑人青少年的性决定和探讨轮椅使用的羞耻感,以及公众对轮椅使用者的反应。

方便样本也用于定性和定量的数据收集。将方便样本用于定量研究数据收集的一个典型例子是调查一门特定课程里的学生,这种课程一般都是由研究者或研究者的朋友教授的课程。这些学生基本上都是在教学环境里"被抓来的"参与者,都是研究者容易找到的。用方便样本收集定性信息的一个典型例子是进行家庭成员的深度访谈。大多数家庭成员,特别是居住在同一户内的那些家庭成员都有可能参加访谈,因为与陌生人相比,他们更可能与调查员建立起信任关系,更重要的是,这些人都还是可以马上进行调查的。

——VoonChin Phua

(沈崇麟译校)

# 谈话分析(Conversation Analysis)

人们互相交谈时,他们不只是在交流想法、信息或知识。在谈话中,如同在所有形式的互动中一样,人们在交谈中做一些事情。他们忙于互相开展社交活动,例如责怪、抱怨、邀请、教导、诉说他们的麻烦等。谈话分析(CA)研究的就是互动对话的参与

者是如何组织他们的言语行为以及非言语行为来进行这种社会行动,或更广泛地说,来处理社会事务的。因此,谈话分析主要是一种分析社会行动的方法(Schegloff,1996)。从方法论上讲,它的目标是发现参与者进行和解释社会行动的那些实践、模式以及一般意义上的方法。

CA起源于哈维·萨克斯(Sacks,1935—1975)1963—1964年在洛杉矶自杀科学研究中心(the Center for the Scientific Study of Suicide, in Los Angeles)的工作。萨克斯对两个方面的问题感兴趣:一个是对中心成员的实践理性方法的民族学兴趣(源自他与哈罗德·加菲尔德的交往);另一个是对互动研究的兴趣(因他读研究生期间受欧文·戈夫曼的启发),他开始分析打给自杀预防中心(the Suicide Prevention Center, SPC)的电话。萨克斯对打进自杀预防中心电话的人们的困境有着高度敏感,开始探究致电者对他们的困难的描述是如何在与SPC的顾问的电话谈话过程中生成的。这促使他去研究通用的谈话的轮流"机制",以及谈话活动管理中的话序模式和结构。萨克斯收集了更广泛的互动语料,包括团体治疗会和普通日常电话的谈话,在与盖尔·杰斐逊和伊曼纽尔的合作之下,他开始铺陈出一幅谈话组织的复杂画面,包括轮次交替、言谈重叠、修复、话题开启和结束、问候、提问、邀请、请求等,以及与这些相关的语序(邻接对)、同意与否定、讲故事和在讲话中加入非语言行为等(Sacks,1992;对谈话分析的各种主题和方法论的全面评论可参见ten Have,1999)。近30年谈话分析的后续研究说明,上述这些互动谈话,以及其他的技术层面在何种程度上是结构化的、社会性地组织起来的资源——或方法——而参与者通过一起交谈开展行动并协调其行动。因此,这些活动是人们建设社会生活的技术基础,或者换个说法,是他们彼此建构其社会性感觉的根基。

CA在很多方面都介于社会学和其他相关学科,特别是语言学和社会心理学之间的交叉地带。当然,CA研究已随着社会语言学、语用学、话语分析等学科的发展一起发展,采用自然状态下发生的互动录音作为基本的数据形式,向一门研究真实言语行为的自然主义的、以观察为基础的科学迈进(Heritage,1984)。尽管CA和其他学科有联系,从方法上来说,在以下几个方面它还是具有社会学的特色的。第一,它关注参与者在谈话轮次中如何相互理解并作出反应,探究的是主体间的社会和互动基础——对社会行动者之间共同的、共享的甚至是"集体"的理解的维护。第二,CA从经验上发展了戈夫曼(Goffman,1983)对社会互动的洞见,即社会互动体现了特别的道德和制度秩序,可以像其他诸如家庭、经济、宗教等社会制度一样来研究。在戈夫曼看来,"互动秩序"就是互动的制度秩序,CA研究的就是让这一制度自身成为一种专题的那些实践。第三,所有层面的语言生产(包括语法的、韵律的、语音的)都可以与人们所从事的那些与他人互动的行动(如问候、邀请、请求)或者活动(如教导、交叉审问、进行医学检查诊断等)关联起来。就此而言,谈话组织是社会行动的基础(Atkinson & Heritage,1984);因此,CA在分析语序的基础上探究行动是如何生产和嵌入的,提供了一种研究如何实现社会行动的方法论。它对于言语行为、非言语行为同样适用,对于把两者结合起来的面对面互动也一样适用(Goodwin,1981)。第四,很显然一种行动(如问候、提问、邀请等)中某位参与者所进行的行动对其他人,也即受动者的反应有所期待;也就是说,他或她可能会期望得到一个问候、对问题的回答、对邀请的接受或是拒绝等。因此,诸如这样的行动对,在CA中称为毗邻语对,它是让行动恰当、负责任的规范性的

框架:行动的规范性特点和与之相关的遵循规范期望的行动的责任,与社会学对社会生活道德秩序的感觉,包括对越轨行为的归因密切相关。第五,CA 将交谈与社会背景相联系。CA 处理背景的方法很特别,部分原因是因为交谈中任何轮次最直接的背景都被当作其所从属的(行动的)顺序,尤其是刚完成的那个轮次。而且,CA 的立场是一个互动的"背景",因此,不可能完全由分析者先验地加以定义;与之相反,参与者用他们设计其交谈的特别方式来展现他们对相关背景的感觉(相互理解,包括每个人对他人有些什么了解;环境;相关的个人信息;他们相关的身份和关系等)。

最后,作为一种问询的方法,CA 也可以像用于日常谈话那样用于制度性的谈话(Drew & Heritage, 1992)。法律、医疗、企业、商务谈判、顾问、教育和公共传播媒体(尤其是新闻访谈)如何管理其制度活动,以及其他的制度性的工作环境的社会世界,就是通过互动交谈来管理的。CA 日益关注这类环境中的参与者如何管理他们各自的制度活动,例如,作为专业人士(医生、律师等)或作为外行(委托人、病人、学生等)。从方法论上讲,CA 研究制度性互动的目标就是揭示那些参与者借此将他们的互动作为一种专门的互动来管理的实践——例如,当作新闻访谈(Clayman & Heritage,2002)、当作咨询、当作法庭上的交叉审问等。在这类研究中,CA 与民族学、社会学直接关联了起来,为研究社会生活的组织和制度现场提供了特别、严格、成果丰富的方法论。

——Paul Drew

(杨可译　沈崇麟校)

## 参考文献


Atkinson, J. M., & Heritage, J. (Eds.). (1984). *Structures of social action: Studies in conversation analysis*. Cambridge, UK: Cambridge University Press.

Clayman, S., & Heritage, J. (2002). *The news interview: Journalists and public figures on the air*. Cambridge, UK: Cambridge University Press.

Drew, P., & Heritage, J. (Eds.). (1992). *Talk at work: Interaction in institutional settings*. Cambridge, UK: Cambridge University Press.

Goffman, E. (1983). The interaction order. *American Sociological Review*, *48*, 1-17.

Goodwin, C. (1981). *Conversational organization: Interaction between speakers and hearers*. New York: Academic Press.

Heritage, J. (1984). *Garfinkel and ethnomethodology*. Cambridge, UK: Polity.

Sacks, H. (1992). *Lectures on conversation* (2 vols.). Oxford, UK: Blackwell.

Schegloff, E. A. (1996). Confirming allusions: Toward an empirical account of action. *American Journal of Sociology*, *2*, 161-216.

ten Have, P. (1999). *Doing conversation analysis*. Thousand Oaks, CA: Sage.


# 相关(Correlation)

一般来讲,相关这一概念与关联【Association】十分接近。它是两个定距或定比量度水平的变量之间的统计关联。

首先,不要把相关和因果效应相混淆。实际上,统计研究只有两个变量的因果效应是不可能的,至少在观察研究中是不可能的。即使在所谓的单因效应,如梅毒螺旋体菌污染和感染梅毒这样的极端例子中,也总

是存在其他起作用、有贡献、有修正作用或起相反效应的变量存在。因果效应的存在只有在多变量环境中才能被细分出来，因而它比相关分析要复杂得多。

让我们把问题限制在两个用 $X$ 和 $Y$ 表示的定距变量，并把因果【Causality】问题置于一边。我们暂且假定一个线性模型。注意到相关系数（经常用 $r$ 来表示）和未标准化【Unstandardized】的回归系数 $b$（计算机输出：$B$）之间的差别是很重要的。后者只不过是回归线的斜率，它是作为由回归线（regression line）和 $x$ 轴形成的角的正切计算的。用收入（$X$）和消费（$Y$）作为变量，我们就有了消费的报价（consumption quote），它是得到一个单位的额外收入之后的人们额外花掉的数量，这样 $X$ 的每一个额外单位的 $Y$ 的变化就是$B=\Delta Y/\Delta X$。在下面我们将会看到，实际上有两个这样的回归系数，而相关系数则是它们的几何均值。

## 相关的五种主要特征

从概率性相关出发，每个相关有五个主要的特征：

### 1.相关的性质

对上面提出的简单相关的计算而言，相关的性质是线性的。这意味着通过散点图，我们将对函数式 $E(Y)=b_0+b_{y1}X_1$ 进行估计。在相关系数中，例如 $r=0.40$ 的背后是一个线性模型。许多研究者只把自己的目光投向那个在 0 和 1 或 0 和-1 之间的一个数上，但却常常遗忘了这个背景模型。在很多时候，他们不自觉地把这种线性模型作为理所当然的。他们似乎没有认识到非线性是经常会发生的。

在联邦德国，天主教徒的百分比和 CDU/CSU（基督教民主联盟/基督教社会联盟）选民百分比之间的相关便是非线性的一个实例。人们可能预期会有一个线性相关：根据一个固定的模式，天主教徒越多，CDU/CSU 的选民就越多。然而这种“越多，越多”的模式似乎只是在有着许多天主教徒的社区才是有效的。而在那些只有少数天主教徒的社区，相关变成了完全都是负的：天主教徒越多，CDU/CSU 的选民就越少。结果，整个散点图展示了一个 U 形的模式。起先它是下降的，而从某个天主教徒的百分比向前，它开始向上。因而某种描述一条抛物线的二次曲线函数，比线性函数显示出了与散点图有着更好的拟合。

有许多其他的描述真实情况的非线性函数，如指数函数、logistic 函数和间断函数等。例如 1970 年代的罗马俱乐部报告，用指数函数来描述上升的环境污染（和时间相关）。logistic 函数则被那些反对罗马俱乐部报告的人用来批评罗马报告持有的无节制的污染可能会达到饱和的水平的观点。

### 2.相关的方向

在处理双变量 $X_1$ 和 $Y$ 时，其中一个是自变量【Independent Variable】，而另一个则是因变量【Dependent Variable】。在收入（$X_1$）和消费（$Y$）的问题中，方向 $X_1\rightarrow Y$ 是明显的。在这里，相关系数和回归系数的区别可能已经被阐明了。回归系数表明 $b_{y1}$ 的方向是 $X_1\rightarrow Y$，而 $b_{1y}$ 的方向是 $Y\rightarrow X_1$。另一方面，可以作为 $b_{y1}$ 和 $b_{1y}$ 的几何均值计算的（等于这两个系数的乘积的平方根）相关系数则合并了这两个方向，因此它是一个无方向的对称系数。

初学者可能会认为，因果分析是回归分析的特权，而相关系数是不适合因果分析的。这种观点固然是可以理解的，但却是不够成熟的。实际上，我们在数学上计算的方向，并不等同于因果的方向。每个以 $Y$ 作为因变量的函数 $Y=aX+b$ 都可变成以 $X$ 作为因变量的逆函数 $X=Y/a-b/a$。在这里，方向是一个相对值。然而，在一块石头被扔进

池塘时，这块落石对扩展的波纹作出了解释，但是与之相反的过程，收缩的波纹可能隐含大量的波纹生成器。它们的内聚性将必须予以解释。不仅如此，我们还必须要证明它们都来自一个中心点（内爆的不可能性）。在这里，方向不是一个相对的值。

诚如前述，因果性不等同于回归系数的统计计算。的确，回归系数也是代表了一种方向，且布莱洛克（Blalock，1972）也在他的因果模型【Causal Model】法中使用了这种方法。然而，就因果性而言，这只不过是一种在统计计算中的相对的方向。休厄尔·赖特（Wright，1934）在他的文章《路径系数法》（*The Method of Path Cofficients*）和早期的文章中也完全承认了这一点。那时他声称已经找到了连接（先验的）因果知识和相关矩阵中的（后验的）经验相关的方法。

3.相关的符号

一个相关既可以是正的（$X_1$越多，$Y$越多），也可以是负的（$X_1$越多，$Y$越少）。在收入和消费这个问题中，二者是一个正相关（符号是+）：收入越高、消费越高。

在20世纪初，社会阶层和孩子数目之间的关系便是一个负相关的例子。较高的社会阶层显示一种被称为社会细管作用（social capillarity）的机制。弗里德里希·恩格斯在他的著作《家庭、私有制和国家的起源》（*Der Ursprung der Familie, des Privateigentums und des Staates*）中，对这个问题作过引人注目的描述。在较高的社会阶层中，只要儿子一生下来，父母就开始实施生育控制，因为这样财产就可以从父亲传给儿子。而在较低社会阶层中，则不存在这样一种机制。一家有10个孩子，是经常的事，有的大家庭孩子甚至多达20个。在这里，符号是极为清楚的，那就是社会阶层越高，孩子数目越少。

4.相关的强度

相关可以是弱的或强的。相关系数为0.90或-0.90，就是强的。相关系数为0.10或-0.10，就是弱的。相关系数为0表示不相关。相关系数为1或-1表示最大的确定性相关。

将向量模型牢记于心可给我们带来很大的方便。向量模型是一个变量的作用和单位一直发生变化的坐标系统：单位是系统的坐标轴，而变量则是形成散点图的点。这些变量被设想为起于原点的向量的终点——故而被称为向量模型。在这样的向量模型中，两个变量之间的相关等同于有两个向量形成的角的余弦。我们知道90°的余弦等于0：那时两个向量互为直角，因而相关系数是0；另一方面，0°的余弦是1：那时两个向量重合，因而达到了最大的相关。

在大多数情况下，一个强的相关系数（在两个变量之间，不是在多变量的情况下）将与一个强的回归系数相一致，但情况并非总是如此。在许多情况中这都是可能的，且与变量的离散程度有关，变量的离散问题将在下面进行介绍。

此外，虽然强的相关系数的确比弱的相关系数更加容易比较显著，但是不显著的强的相关系数这样的情况也时有发生。不仅如此，一个显著的相关系数也可能是弱的。由此可见，一个相关（相关或回归系数的）的强度是不可以这样来判断的，因为它必须和其他标准合起来加以研究，首先应该考虑的问题便是接着要讨论的显著性问题。

5.统计可能性的概推

根据随机样本计算的相关可能是显著的也可能是不显著的（即它是否可以被推广到那个我们从中选取样本的总体）。在这里，我们必须站在一系列后现代思想的反面，而采取一个传统的严格的观点：只有那些在统计上可以推广的（因而是统计上显著

的)相关才适合进行解释。不显著的相关只是随机性的结果,因而应当被认为在总体中是不存在的。

因为样本太小(例如 $n=5$),因此一个强的(例如 $r=0.90$)而不显著的相关系数是经不起检验的,因此这也为任何社会研究者所拒绝。一个有识别能力的头脑将拒绝在一个足够大的(例如样本量 $n=5\ 000$)样本中的,一个弱的(例如 $r=0.10$),但仍然显著的系数,而它之所以显著只是因为这个样本量。取一个足够大的样本,我们可以用统计值来证明任何东西,因为它终究是会出现的。然而这种观点只有部分是正确的,在处理大样本时,相关或微小的变化,如生育力的变化可能对政策制订者来讲都是极有意义的(在重要性上而言)。任何人口学家都知道这样的过程是根本不可逆的。此外,即使对大批量(的样本)而言,相关仍然有可能在统计上是不显著的。因此,我们不能假装这种结果仅仅是因为样本量的缘故。

应当承认,在两个极端之间仍然存在着一个灰色地带:小样本的显著结果和大样本的不显著结果。在后一种情况下我们确切地知道相关是不显著的。在前一种情况下,我们能确定一个较大的样本调查将会产生一个显著的结果。然而,一般来讲,当然是指灰色地带而言,非统计的标准,如与社会关联的标准将总是会有助于相关结果的解释。

迄今为止,我们已经了解了相关的某些性能:性质、方向、符号、强度和显著性。正如我们已经提到的一样,我们不应该为数字代表的相关系数所蒙蔽,事实上,所有这些特性都应当在考虑之列。从这个意义上讲,在重要的元分析【Meta-Analyses】中,来自不同研究的相关系数都被混在了一起,我们不应该轻信它们。事实上,在每一个不同的研究中,我们都应该提出许多疑问:我们是否在处理一个线性相关?如果是,回归系数与相关系数是否有差别,截距又有多大?变量是否已经标准化?相关是否显著?置信区间有多大?第一类误差【Type Ⅰ Error】$\alpha$ 的概率是多少?样本量有多大?第二类误差【Type Ⅱ Error】$\beta$ 和检验力的概率是多少?

## 变量的离散度:标准化系数

相关系数和回归系数的差别实际上也涉及变量的离散度。确实,我们已经提到过强的相关系数并非总是刚巧会有一个强的回归系数,而这与变量的离散度有联系。这一点可以从系数的公式推导出来:$r_{y1}$(或 $r_{1y}$)是协方差除以方差的几何均值,$b_{y1}$ 是协方差除以 $X_1$ 的方差,$b_{1y}$ 是协方差除以 $Y$ 的方差。在等离散度的情形中,所有这三个系数将是相等的。$Y$ 的离散比较大会使相关系数 $b_{1y}$ 下降,而非 $b_{y1}$ 下降。

有关这一问题,有人可能想知道标准化系数($\beta$)是否是首选(标准化的方法是系数值减去均值,再除以标准差【Standard Deviation】)。除以标准差是一个非常激烈的操作,它会使变量的方差被人为地置为1,以便因此而得到可比性。但有人认为,教育中类似的差别,在存在大量教育不平等的社会,与那些有着较低程度同样差别的社会相比,有着不同的意义,因而他们提倡使用非标准化的系数。

所有这一切并非只是抽象的思维,它的确发生在社会的特定发展过程。这方面的例子包括有效的生育控制过程、就业不足、提升政治的公共舆论和环保意识等。从所有这些例子中我们了解到,社会是在它的这些特征的离散中引导变迁的。

## 相关系数的计算

我们想用计算教育($X$)和收入($Y$)之间相关的小例子来阐述这种离散度的重要性。计算公式是 $r_{xy} = [\sum(X-\overline{X})(Y-$

$\bar{Y})]/[\sum(X-\bar{X})^2\sum(Y-\bar{Y})^2]^{1/2}$。

让我们来取一个公民上学很少的国家，那些刚好受过教育的人恰好都是高收入者，而其余的人则都很穷。令数据矩阵如下面所列。教育用一个 1~9 的尺度来量度，而收入则以千欧元来量度。

*X* 1 1 1 2 2 3 3 3 4 4 6 6 6 7 7 8 8 8 9 9

*Y* 2 4 5 3 5 3 4 6 4 5 5 7 8 6 8 6 7 9 7 8

散点图如图 1 所示。

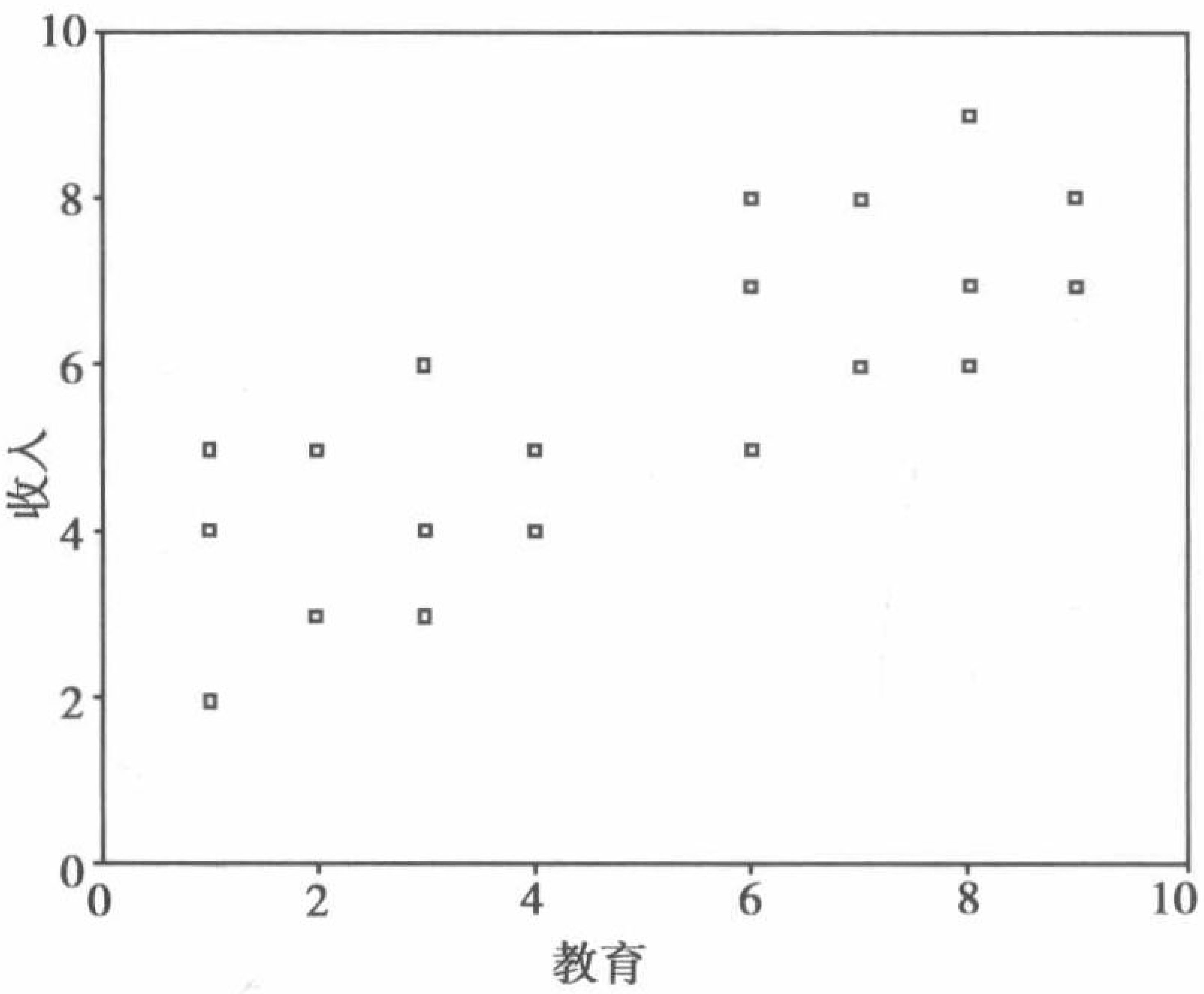



图 1

教育和收入的标准差分别为 2.81 和 1.93。

相关系数是 0.798。该模型是线性的，相关系数是对称的，关系是正的，强度是很强的，而系数也是在统计上显著的($p<0.001$)。现在我们假设在这个国家，穷人中更有上进心的人希望自己的孩子能进学校读书，而国家迫于这种压力，开始资助教育。大量的人到中学和大学读书，随着教育水平的提高，人的能力水平、平均收入水平也随之上升，这时收入的变差下降。但是，一旦资助教育变得普遍，那些读过更多书的人将再也不能从学校教育中获利，致使收入和教育的相关减弱。

数据矩阵如下所述：

*X* 6 6 6 6 6 6 6 7 7 7 7 8 8 8 8 8 8 9 9 9 9

*Y* 5 7 8 5 7 8 6 8 6 8 6 7 9 6 7 9 7 8 7 8

散点图如图 2 所示。

标准差变得小得多，教育和收入分别为 1.14 和 1.17。相关系数从 0.798 下降到了 0.285。关系是正的，但变小了，因而在统计上不再显著。

人们也可以假设，在这个国家，教育的变差仍然相同，因为只有一些公民进学校读书，但是收入的变差却下降了，因为存在着许多不充分的就业者(即高学历低收入者)。这将导致如下的数据矩阵：

*X* 1 1 1 2 2 3 3 3 4 4 6 6 6 7 7 8 8 8 9 9

*Y* 2 4 5 3 5 3 4 6 4 5 2 4 5 3 5 3 4 6 4 5

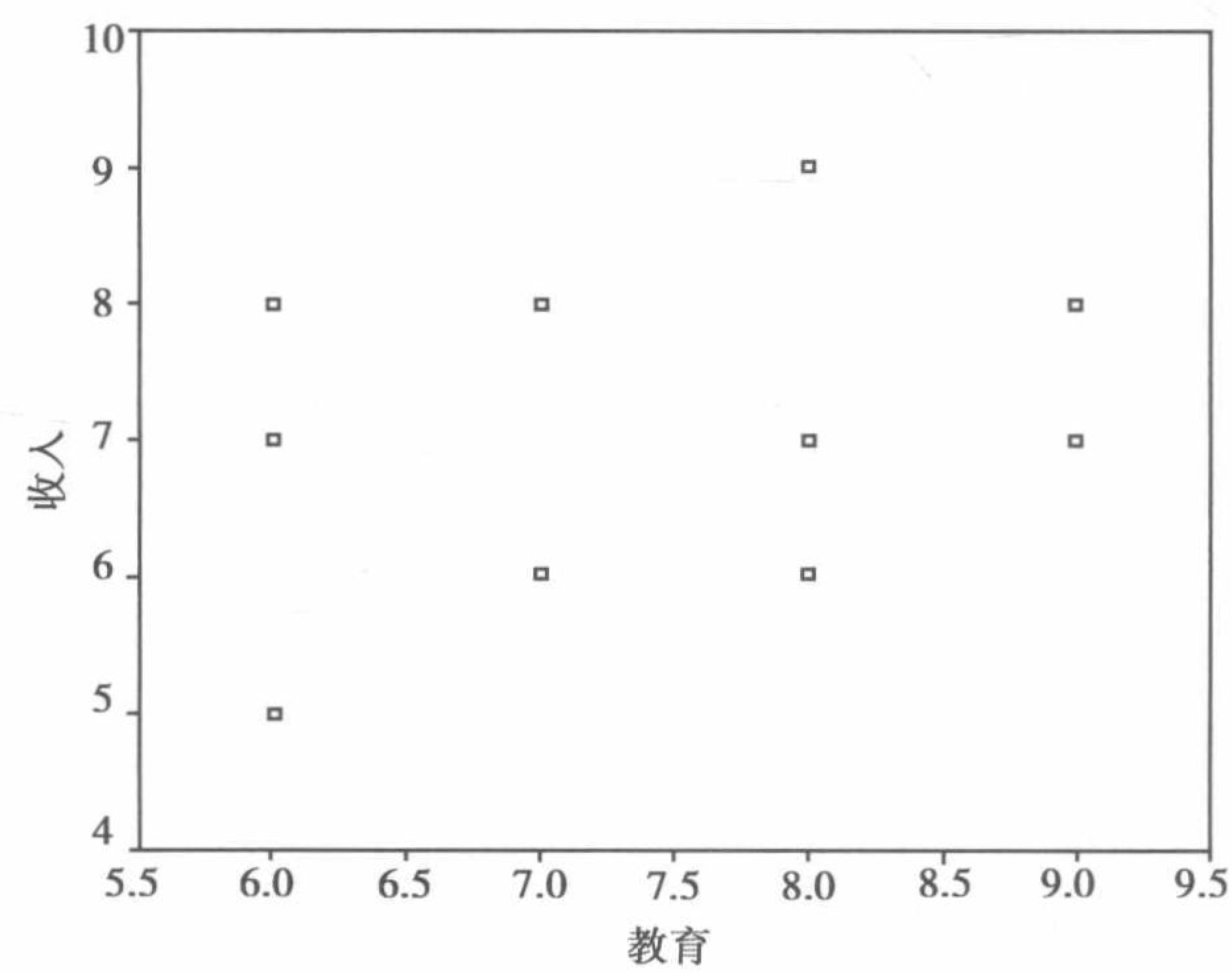



图 2

散点图如图 3 所示。

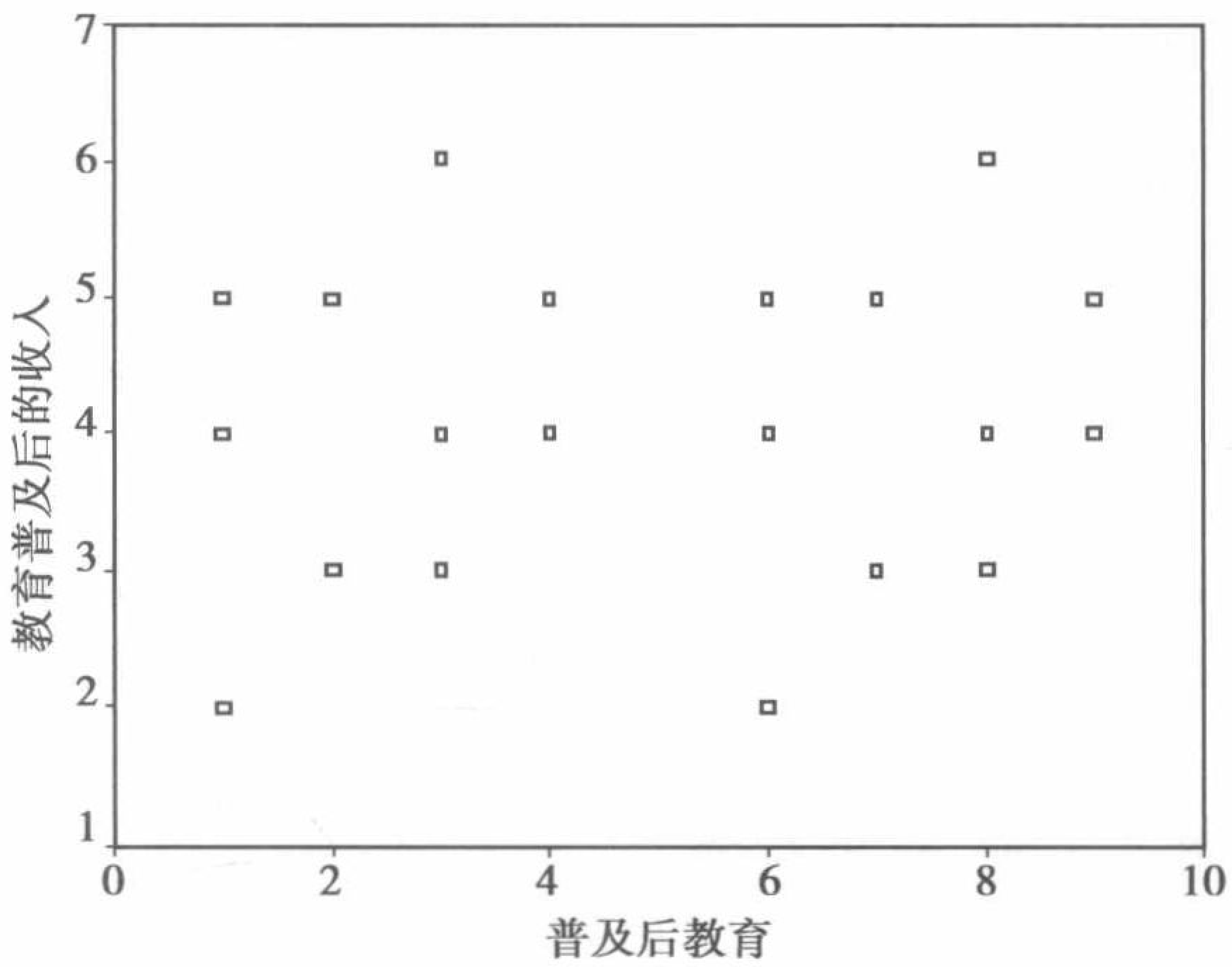



图 3

教育的标准差仍然没有变(2.81),但收入的标准差却变小了(1.17)。现在相关系数只有 0.116。关系仍然是正的,但却变得更小了,而且是不显著的。

——Jacques Tacq

(沈崇麟译校)

## 参考文献


Blalock, H. M . ( 1972 ). *Social statistics*. Tokyo: McGraw-Hill Kogakusha.

Brownlee, K. (1965).*Statistical theory and methodology in science and engineering*.New York:John Wiley.

Hays, W. L. (1972). *Statistics for the social sciences*. New York: Holt.

Kendall, M., & Stuart, A. (1969). *The advanced theory of statistics*.London: Griffin.

Tacq, J.J.A. (1997). *Multivariate analysis techniques in social science research: From problem to analysis*. London:Sage.


# 对应分析(Correspondence Analysis)

对应分析是一种洞察,而非量度双向交叉表中的关联【Association】的方法。例如,我们来看一下表 1 中的交叉表。该表缘于 1994 年实施的一个题为"家庭和变化中的性别角色(Family and Changing Gender Roles)"的国际社会调查项目(ISSP)。这张表显示,在 24 个不同的国家,回答人是如何考虑家庭的理想小孩数为 0,1,2,3,4,5 或更多个这个问题的。每个国家的回答的百分比都列在了括弧中。例如,在澳大利亚,1 529个回答人中的 833 个或 54.5%的回答人认为理性的孩子数是两个。

这张 24 × 6 的表通常的 Pearson 卡方统计量的值为 6 734,而 Cramér 的 $V$ 是 0.206,二者都是高度显著的。实际上,任何关联量度【Measure of Association】,即使有一个较低的值,对一个有着如此高的频数的表而言,也几乎总是高度显著的。很清楚,各个国家之间都存在着显著的差异,但这些差异是什么呢?我们从这张表上看到,美国和加拿大颇为相似,但我们如何才能将它们在这一问题上的位置和欧洲国家作比较。对应研究便可以回答这个问题,因为它的目的是以简洁的图示方式来显示国家之间的

异同。

表 1 的对应分析显示在图 1 中。这是一张行和列点的二维的位置“地图”。它显示列联表的主要特性。图中代表列的点形成了一种弓形纹，始于底部的右方，终止于底部的左方。这表明这些国家在这一问题上的态度是单调排列的，即从那些想要比较少的孩子（如民主德国、保加利亚和捷克）到想要较多的孩子（如爱尔兰、以色列和菲律宾）。水平轴或第一主轴捕捉这些国家的基本比例尺，而纵向的第二主轴，将极端的类别与中间的类别进行比较。诸如这样的弓形纹经常会在对应分析的地图中遇到。大多数高度发达的国家都围绕在地图的中央，中央表示平均，但在他们之中存在着令人感兴趣的差别。例如，美国与其他围绕中央的国家相比，位于第二根轴的往下部位，主要是因为美国回答想要“4 个孩子的”频数相对较高。与之相反，日本则高居第二根轴的上方，因为其回答希望要“3 个孩子”的频数很高。

**表 1　家庭理想孩子数：24 个国家的答案频数（括号中的数字为百分比）**

| 国家 | 0 | 1 | 2 | 3 | 4 | ≥5 | 合计 |
|---|---|---|---|---|---|---|---|
| 澳大利亚 | 4(0.3) | 19(1.2) | 833(54.5) | 448(29.3) | 203(13.3) | 22(1.4) | 1 529 |
| 联邦德国 | 18(0.8) | 116(5.2) | 1 551(69.5) | 441(19.8) | 79(3.6) | 26(1.2) | 2 231 |
| 民主德国 | 6(0.6) | 108(10.1) | 831(77.7) | 115(10.7) | 10(3.6) | 0(0.0) | 1 070 |
| 英国 | 5(0.6) | 16(1.8) | 679(74.9) | 145(16.0) | 55(6.1) | 6(0.7) | 906 |
| 北爱尔兰 | 0(0.0) | 7(1.2) | 271(45.5) | 157(26.4) | 142(23.9) | 18(3.0) | 595 |
| 美国 | 11(0.8) | 34(2.6) | 783(60.0) | 295(22.6) | 159(12.2) | 22(1.7) | 1 304 |
| 奥地利 | 5(0.5) | 51(5.4) | 667(70.4) | 195(20.6) | 28(3.0) | 2(0.2) | 948 |
| 匈牙利 | 8(0.6) | 70(4.8) | 839(57.7) | 485(33.4) | 41(2.8) | 11(0.8) | 1 454 |
| 意大利 | 4(0.4) | 46(4.5) | 678(67.1) | 254(25.1) | 22(2.2) | 7(0.7) | 1 011 |
| 爱尔兰 | 2(0.2) | 7(0.8) | 276(31.3) | 278(31.5) | 270(30.6) | 49(5.6) | 882 |
| 荷兰 | 41(2.1) | 42(2.1) | 1 050(53.4) | 575(29.2) | 199(10.1) | 61(3.1) | 1 968 |
| 挪威 | 3(0.2) | 10(0.5) | 939(48.2) | 824(42.3) | 145(7.4) | 29(1.5) | 1 950 |
| 瑞士 | 0(0.0) | 8(0.7) | 741(63.9) | 329(28.4) | 65(5.6) | 16(1.4) | 1 159 |
| 捷克 | 7(0.7) | 109(10.7) | 688(67.3) | 192(18.8) | 16(1.6) | 11(1.1) | 1 023 |
| 斯洛文尼亚 | 9(0.9) | 40(3.9) | 601(59.2) | 311(30.6) | 45(4.4) | 9(0.9) | 1 015 |
| 波兰 | 0(0.0) | 27(1.9) | 756(54.1) | 483(34.6) | 99(7.1) | 32(2.3) | 1 397 |
| 比利时 | 55(4.9) | 80(7.1) | 742(65.9) | 216(19.2) | 29(2.6) | 4(0.4) | 1 126 |
| 俄罗斯 | 0(0.0) | 188(9.6) | 1 219(62.5) | 463(23.7) | 50(2.6) | 30(1.5) | 1 950 |
| 新西兰 | 8(0.9) | 11(1.2) | 497(55.7) | 258(28.9) | 106(11.9) | 12(2.3) | 892 |
| 加拿大 | 5(0.4) | 25(2.2) | 674(59.3) | 308(27.1) | 107(9.4) | 17(1.5) | 1 136 |
| 菲律宾 | 0(0.0) | 10(0.8) | 229(19.1) | 561(46.8) | 282(23.5) | 117(9.8) | 1 199 |

续表

| 国家 | 0 | 1 | 2 | 3 | 4 | ≥5 | 合计 |
|---|---|---|---|---|---|---|---|
| 以色列 | 1(0.1) | 9(0.7) | 204(16.7) | 525(42.9) | 384(31.4) | 100(8.2) | 1 223 |
| 日本 | 0(0.0) | 9(0.7) | 483(37.2) | 737(56.7) | 56(4.3) | 15(1.2) | 1 300 |
| 西班牙 | 0(0.0) | 140(6.0) | 1 441(61.7) | 593(25.4) | 108(4.6) | 55(2.4) | 2 337 |
| 合计 | 192(0.6) | 1 182(3.7) | 17 672(55.9) | 9 188(29.1) | 2 700(8.5) | 671(2.1) | 31 605 |

注:表 1 的 ISSP 的数据来自德国科隆的 *Zentralarchiv für Empirische Sozialforschung*(德国科隆大学实证社会研究中心)。

每个国家都可以根据它的图像或一组相对频数(在表 1 中以百分比列出)在地图中找到它的位置。表的中央代表平均图像,那就是所有国家的合计0.6%,3.7%,55.9%,29.1%,8.5%和 2.1%(见表 1 的最后一行)。如果一个国家的某一类别的百分比高于相应的平均值,那么它就将被吸引向某一类。这种方法也可被定义为数值型变量【Metric Variable】的变差的多维度尺度。在该尺度中国家之间的距离以所谓的卡方距离来量度(Benzécri, 1973),而每个国家的权重则与其样本量成比例。

在构建这种地图时,主要的问题是如何从几种不同的可供选择的尺度中确定选择一种作为轴的尺度。图 1 这样的图称为对称地图,因为它表示行和列是以同样的方式来使用各自的主坐标的,所以在每一组点在地图中都着相同的离散。在每个国家都位于答案类别点的加权平均的位置时,可能会出现一张不对称的地图,后者以标准坐标来表示(Greenacre, 1993)。数据表中的方差是用惯量(inertia)来量度的,它等于 Pearson 卡方统计值除以表的总和。显示的质量则以惯量在每一主轴上所占的百分比量度,它与主成分分析【Principal Conponents Analysis】类似。在图 1 中,两根轴解释了惯量的组合 70.5%+14.7%=85.2%,剩下未解释的 14.8%位于沿主轴的更远处。

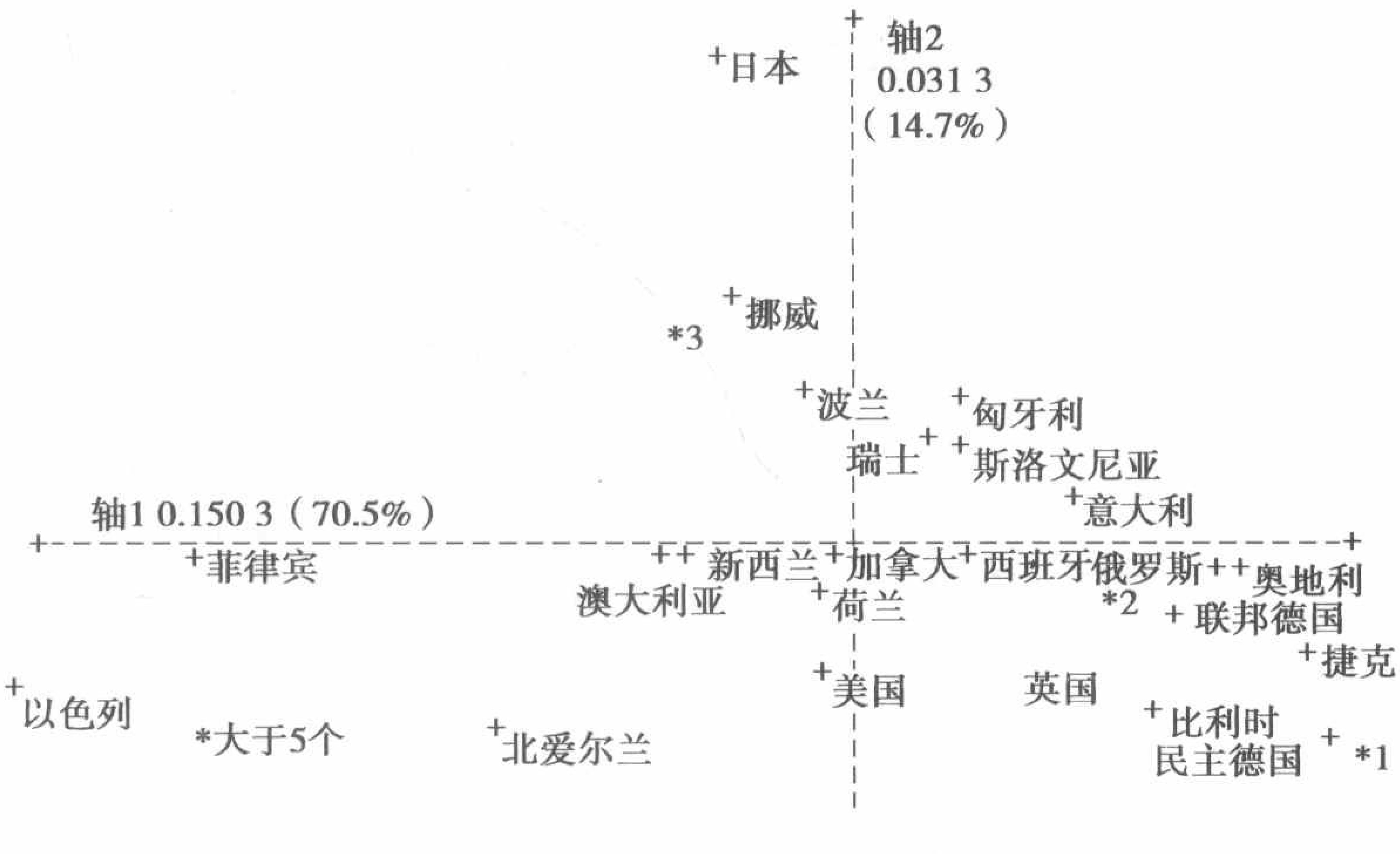



图 1　表 1 的对应分析:双维度对称地图

尽管对应分析主要用于列联表【Contingency Tables】,但它也可以用于多变量类别数据的直观解释模式(多变量对应分析)、参数选择、排序和评级。而要达到这样

的目的,则需要在应用对应分析之前,对原始数据进行恰当的转换。

——Michael Greenacre

(沈崇麟译校)

## 参考文献

Benzécri, J.-P. (1973). *Analyse des Données. Tome 2: Analyse des Correspondances* [Data analysis: Vol. 2. Correspondence analysis]. Paris: Dunod.

Greenacre, M. J. (1984). *Theory and applications of correspondence analysis*. London: Academic Press.

Greenacre, M. J. (1993). *Correspondence analysis in practice*. London: Academic Press.

Greenacre, M.J., & Blasius, J. (1994). *Correspondence analysis in the social sciences*. London: Academic Press.

# 平衡(Counterbalancing)

在一个重复测量设计【Repeated-Measures Design】中,如果参与者都要经受所有的处理条件,那么我们就需要考虑延滞效应【Carryover Effects】和次序效应【Order Effects】问题。平衡指的是让参与者经受不同顺序的处理条件,以保证诸如这样的延滞效应和顺序效应相等地落在实验条件上。

平衡可以是完全的和不完全的。拿完全的平衡来说,每种可能的处理顺序被采用的机会均相等。因此,在 $k$ 次处理中,会有 $k!$ 个不同的顺序。如表 1 所示,对三种处理条件(A, B 和 C),每种处理在每个位置出现的次数都一样。不仅如此,每个处理的处理次数也与前面的每一个其他处理的相等(见表 1)。

**表 1**

| 顺序 | 1 号位置 | 2 号位置 | 3 号位置 |
|---|---|---|---|
| 1 | A | B | C |
| 2 | A | C | B |
| 3 | B | A | C |
| 4 | B | C | A |
| 5 | C | A | B |
| 6 | C | B | A |

理解平衡的好处最好的办法是,思考在上面的实验中如果没有平衡会发生什么。假设所有的参与者接受处理的顺序都相同(A,B,C),而且实验发现,在条件 C 下的平均绩效显著优于条件 A 下的。这个结果无法得到确凿无疑的解释,因为处理和位置混杂【Confounding】在一起了。也就是说,差异可能源自处理(A 对 C)或者来源于一种练习效应(1 号位置对 3 号位置),在此情况下,不论处理是什么,参与者的分数都将会与时俱进。

平衡接受处理的顺序可以排除处理和顺序之间的混淆。不过,这么做也会增大分析中的残差项,因为每种处理条件下任何系统的顺序效应都会增加分数的变异性。顺序效应可以进行估计并从残差中除去,已有研究对此进行了详细介绍(Keppel, 1991, pp.361-373)。

完全的平衡很理想,因为它不仅能保证每种处理出现在每个位置上的次数相同,也能保证每种处理都能以相等的次数出现于所有位于前面的位置(例如,1 号位置、2 号位置等)。不过,如果要做的处理很多,完全的平衡可能不现实。在那种情况下,应退而求其次,采用不完全的平衡。

有一种有用的不完全平衡方法,称为平衡连字的拉丁方【Latin Square】。使用这种方法,如果处理的数目($k$)为偶数时,那么

只需要 $k$ 个顺序。如果处理的数量是奇数，那么就需要 $2 \times k$ 个顺序排列。

每个顺序都是按以下方法建构的。把各种处理条件排成一行(ABCDE)。第一个顺序就是第一个、第二个、最后一个、第三个、倒数第二个、第四个,以此类推。生成第二个顺序时,首先要重新排列你的条件行,即把第一个处理放到最后,其他的条件还保留原来的顺序(BCDEA)。接下来按照以前同样的方法选择你的顺序(第一个、第二个、最后一个等)。一直用同样的方式重新排列和选择,直到得到 $k$ 个顺序。当 $k$ 是偶数时,你在确定了 $k$ 个顺序之后就可以结束了。

当 $k=5$ 时,用这个程序可以得到左边的五个顺序(见表 2)。每个处理在每个位置上出现的频率相等,这很重要。不过,因为 $k$ 是奇数,所以并非每个处理之前都已出现过每个其他的处理。例如,如表 2 中的顺序 1 至顺序 5,B 的前面没有出现过 C 或 E。因此,当 $k$ 是奇数时,有必要像顺序 6 到顺序 10 那样把每个顺序都进行一下镜像排列(逆序排列)。你可以从这个例子中看出,平衡连字的方法保证了每个条件在每个位置上出现的机会相同,而且每个条件前面每个其他条件出现的频次也都相等。

表 2

| 顺序 | $P_1$ | $P_2$ | $P_3$ | $P_4$ | $P_5$ | 顺序 | $P_1$ | $P_2$ | $P_3$ | $P_4$ | $P_5$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | A | B | E | C | D | 6 | D | C | E | B | A |
| 2 | B | C | A | D | E | 7 | E | D | A | C | B |
| 3 | C | D | B | E | A | 8 | A | E | B | D | C |
| 4 | D | E | C | A | B | 9 | B | A | C | E | D |
| 5 | E | A | D | B | C | 10 | C | B | D | A | E |

——Hugh J. Foley

(杨可译　沈崇麟校)

## 参考文献


Keppel, G. (1991). *Design and analysis: A researcher's handbook* (3rd ed.). Englewood Cliffs, NJ: Prentice Hall.

Maxwell, S. E., & Delaney, H. D. (2000). *Designing experiments and analyzing data: A model comparison perspective*. Mahwah, NJ: Lawrence Erlbaum.


# 反事实(Counterfactual)

反事实论断是一种其前提条件不存在的,但如果我们得到了这个前提条件,那么它产生的后果所描述的世界将会是什么样的论断。反事实采用虚拟条件的形式:"如果 $P$ 存在,那么我们将会得到 $Q$。"为了理解和评估这样一个论述,我们需要考虑假设前提条件能够获得时,世界将会是什么样的。例如,"如果风速没有达到每小时 50 英里,桥就不会坍塌"或者"如果安全理事会采取了行动,战争将会被避免"。我们可以提出

两种类型的反事实条件问题：这个论述的意义是什么？以及我们如何确定它是真的还是假的？反事实条件问题不能像真值函数条件问题那样进行评估，因为带有虚假前提的真值函数条件问题事实上仍然是真实的（例如，"如果 $P$ 所以 $Q$"，等价于"非 $P$ 或非 $Q$"）。所以，如果反事实要在严密的思想中发挥作用，就非常有必要对反事实问题的真实条件提供符合逻辑的分析。

反事实推论和因果推理之间有着非常紧密的关系。如果我们断言"$P$ 导致 $Q$（在情境 $C_i$ 下）"，这意味着我们的断言如下："如果 $P$ 没有发生（在情境 $C_i$ 下），那么 $Q$ 也不会发生。"所以一个因果判断意味着一组反事实判断。与此相对应，一个反事实判断通常会从一个或多个，将前提情况导致的结果传达给世界的因果过程得到支持。当我们说，如果风速没有达到每小时 50 英里，塔科马市的纽约海湾（Tacoma Narrows）大桥就不会坍塌时，我们依靠的是关于大桥的结构力学和风在造成该结果方面的作用的因果性理论。

我们如何才能将一个真值赋予反事实陈述呢？最有条理的答案是诉诸因果关系和因果定律。如果我们相信对于 $Q$ 的发生给出了真实的因果性分析，而如果 $P$ 是使得 $Q$ 出现的充足条件中的一个必要部分，那么我们就可以认为，如果 $P$ 发生，$Q$ 就会发生。大卫·刘易斯（Lewis，1973）用可能的世界（可能世界符号学）分析了真实条件和反事实的逻辑。他把反事实解释为一种有关在同样的自然法则支配的其他可能的世界中，事情是如何发生的陈述。大致来说，"在每个与现存世界贴切相似，但风速没有达到每小时 50 英里的可能世界中，大桥就不会坍塌"。构成世界间"贴切相似"的事物则被用"遵循相同的自然法则"来解释。在这种路数中，我们可以将"如果 $P$ 发生了，$Q$ 会发生"的反事实理解为一个遵循这些原则的陈述："$P$ 和｛自然法则｝引起了 $Q$。"这个建构将"物理的必然性"概念引进了有关反事实的陈述：给定 $P$，$Q$ 就会发生，这是物理的必然性。

——Daniel Little

（郑晓娟译　沈崇麟校）

## 参考文献

Lewis, D. K. (1973). *Counterfactuals*. Cambridge, MA: Harvard University Press.

# 协方差（Covariance）

协方差是两个随机变量之间的关联的基本量度。在这里将要依据双变量正态分布【Normal Distribution】来讨论协方差概念。

首先，让我们来看一下两个代表阅读能力量度和数学能力量度的随机变量【Random Variable】$X$ 和 $Y$。我们还假定两个变量服从双变量正态分布。定义 $E(X)$ 和 $E(Y)$ 为随机变量 $X$ 和 $Y$ 的期望。那么 $X$ 和 $Y$ 的协方差 $\mathrm{Cov}(X,Y)$ 则可表示为

$$\mathrm{Cov}(X, Y) = E[(X-\mu_x)(Y-\mu_y)] \quad (1)$$

式中的 $\mu_x = E(X)$，而 $\mu_y = E(Y)$。总体协方差的无偏估计量的计算公式则为

$$\frac{\sum_{i=1}^{N}(X_i-\bar{X})(Y_i-\bar{Y})}{N-1} \quad (2)$$

式中的 $N$ 是样本量,$\overline{X}$ 是 $X$ 的样本均值,$\overline{Y}$ 是 $Y$ 的样本均值。

回到公式(1),双变量正态分布的均值包含在均值向量 $\mu$ 中,而 $Y$ 则包含在协方差矩阵 $\sum$ 中。协方差矩阵是一个对称矩阵,其对角线是 $X$ 和 $Y$ 的方差,在对角线之外的是 $X$ 和 $Y$ 的协方差。

协方差表示两个随机变量 $X$ 和 $Y$ 之间的线性关系;在 $X-\mu_x$ 和 $Y-\mu_y$ 符号相同时,协方差的值就可能为正;而在 $X-\mu_x$ 和 $Y-\mu_y$ 符号相反时,则协方差就可能有负值。然而不论协方差的符号是什么,它的值都是无意义的,因为它取决于方差,也即取决于随机变量本身的度量。为了缓解这一问题,相关【Correlation】系数将协方差标准化为有均值零和方差 1,因而将随机变量置于一个共同的度量之上。这样将会生成一个-1~+1 的值域,零表示线性关系不存在的线性关联量度。

因为协方差是一个统计量,因而就会有一个抽样分布【Sampling Distribution】。设想从一个具有双变量正态分布特征的总体中反复抽取两个随机变量。对于每一次抽取而言,我们都可以计算两个变量的均值、方差和协方差。均值的抽样分布已知服从正态分布、方差的抽样分布已知服从卡方【Chi-Square】分布。至于协方差矩阵,其矩阵的抽样分布则称为 Wishart 分布。Wishart 分布是一种类似于卡方分布的多变量分布(参见 Timm, 1975)。令人感兴趣的是,用于因子分析【Factor Analysis】和结构方程模型【Structural Equation Modeling】的最大似然【Maximum Likelihood】拟合函数可以从 Wishart 分布中导出。

在定义了两个变量之间的协方差和它的抽样分布之后,我们再来考虑一下它的应用问题。诚如前述,协方差给我们提供了一种两个随机变量之间的线性关联的量度。还是诚如前述,更为普遍的用法是将协方差转换成易于解释的相关系数。不仅如此,协方差——特别是协方差矩阵,还是其他方法的重要量度。许多多变量统计方法,都是依据以更为基本的参数表达的协变量来定义的,如回归【Regression】分析、路径分析【Path Analysis】和结构方程建模等方法,都可作为变量的协方差模型。

——David W. Kaplan

(沈崇麟译校)

## 参考文献

Timm, N. H. (1975). *Multivariate analysis with applications in education and psychology*. Monterey, CA: Brooks/Cole.

# 协方差结构(Covariance Structure)

协方差一词用来标明一大类设法用较少参数来代表一组变量的方差和协方差的统计模型。其前提是样本方差和协方差足以进行模型的参数估计。均值结构一词用来表示一类用少数几个参数代表的样本数据的均值的模型。协方差结构的一般问题始见于柏罗文的著作(Browne,1982)。

在协方差结构名下的模型包括,但不限于回归分析、路径分析和结构方程模型,模型的一般形式可写作:

$$\sum = \sum(\Theta) \qquad (1)$$

式中的 $\sum$ 是 $p \times p$ 的总体协方差矩阵，$p$ 是变量的数目，而 $\Theta$ 则是 $q$ 维的模型参数向量，且 $q \leqslant p$。方程(1)中的函数表明，协方差矩阵可以被表示为一个参数向量的一个矩阵值的函数。

诚如前述，方程(1)中的协方差结构模型包括了一些常用的统计程序。这一模型要求模型的参数可用协方差矩阵的元素来表示。我们来看一个简单的双变量线性回归模型的例子。该模型可表示为

$$y = x\beta + u \tag{2}$$

式中的 $y$ 和 $x$ 分别为结果和回归量变量，$\beta$ 是回归系数，而 $u$ 则是回归的干扰项。方程(2)可看成是回归系数 $\beta$ 可按下式

$$\hat{\beta} = \frac{\mathrm{Cov}(x, y)}{\mathrm{Var}(x)} \tag{3}$$

估计时的协方差结构。式中的 $\mathrm{Cov}(x, y)$ 是 $x$ 和 $y$ 的协方差，而 $\mathrm{Var}(x)$ 是 $x$ 的方差。同样，干扰项的方差 $u$(用 $\sigma_u^2$ 表示)可用 $\sum$ 的元素求出。我们看到的那个回归分析是协方差结构模型的一个特例。

再举一个协方差结构代表的普遍性的例子，我们来看一下因子分析模型，它可写作如下的线性形式：

$$\mathbf{y} = \mathbf{\Lambda\eta} + \boldsymbol{\varsigma} \tag{4}$$

式中的 $\mathbf{y}$ 是一个变量的 $p \times 1$ 向量，$\Lambda$ 是一个因子负荷的 $p \times k$ 矩阵，$\boldsymbol{\eta}$ 是一个公共因子的 $k \times 1$ 向量，而 $\varsigma$ 则是一个唯一性的 $p \times 1$ 向量(由量度误差和特殊误差组成)。假定因子的均值和唯一性的均值为零，且因子的协方差和唯一协方差也为零，那么 $y$ 的协方差矩阵可表示为

$$\sum\nolimits_y = \mathbf{\Lambda\Phi\Lambda}' + \mathbf{\Psi} \tag{5}$$

式中的 $\mathbf{\Phi}$ 是因子之间的协方差矩阵，而 $\mathbf{\Psi}$ 则是唯一方差的对角矩阵。由方程(5)可知，协方差矩阵是以因子分析模型的基本参数表达的，因此，由定义可知，它是一种协方差结构。最后我们可能有兴趣看到，结构方程模型是如何拟合进广义协方差结构框架的。结构方程系统可被写作：

$$\mathbf{y} = \boldsymbol{\alpha} + \mathbf{By} + \mathbf{\Gamma x} + \boldsymbol{\varsigma} \tag{6}$$

式中的 $\mathbf{y}$ 是一个观察的内生变量的 $p \times 1$ 向量；$\mathbf{x}$ 是一个观察的外生变量的 $q \times 1$ 向量；$\alpha$ 是一个结构截距的 $p \times 1$ 向量，$\mathbf{B}$ 是内生变量彼此联系的 $p \times p$ 系数矩阵，$\mathbf{\Gamma}$ 是 $p \times p$ 的联系内外变量和外生变量的系数矩阵。而 $\boldsymbol{\varsigma}$ 则是一个干扰项的 $p \times 1$ 向量，其中 $\mathrm{Var}(\boldsymbol{\varsigma}) = \boldsymbol{\varphi}$ 是干扰项的 $p \times p$ 协方差矩阵，$\mathrm{Var}(\cdot)$ 是方差算子。最后令 $\mathrm{Var}(\mathbf{x}) = \mathbf{\Phi}$ 为外生变量的 $q \times q$ 协方差矩阵。

方程(6)中描述的系统称为模型的结构形式。它便于重写结构模型，这样内生变量便可在方程的一侧，而外生变量则可在方程的另一侧。方程(6)可被重写为

$$(\mathbf{I} - \mathbf{B})\mathbf{y} = \boldsymbol{\alpha} + \mathbf{\Gamma x} + \boldsymbol{\varsigma} \tag{7}$$

假定 $(\mathbf{I} - \mathbf{B})$ 是非奇异的，那么它的倒数存在，方程(7)便可被写为

$$\mathbf{y} = (\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\alpha} + (\mathbf{I} - \mathbf{B})^{-1}\mathbf{\Gamma x} + (\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\varsigma} = \mathbf{\Pi}_0 + \mathbf{\Pi}_1 \boldsymbol{x} + \boldsymbol{\varsigma}^* \tag{8}$$

式中的 $\mathbf{\Pi}_0$ 是截距向量的简约式，$\mathbf{\Pi}_1$ 是斜率向量的简约式，而 $\boldsymbol{\varsigma}^*$ 则是有 $\mathrm{Var}(\boldsymbol{\varsigma}^*) = \mathbf{\Psi}^*$ 的干扰向量的简约式。这种设定称为模型的简约式。注意，方程(8)是一个 $\mathbf{y}$ 在 $\mathbf{x}$ 上的简明的多元回归，因此这种模型也可归入广义的协方差结构式。

接下来，令 $\Theta$ 为一个含模型的结构参数的向量，这样在本例中，$\Theta = (\mathbf{B}, \mathbf{\Gamma}, \mathbf{\Psi}, \mathbf{\Phi})$。不仅如此，令 $E(\mathbf{x}) = \boldsymbol{\mu}_x$ 为 $\mathbf{x}$ 均值的向量，$\mathrm{Var}(\mathbf{x}) = E(\mathbf{x}'\mathbf{x}) = \mathbf{\Phi}$ 和 $E(\boldsymbol{\varsigma}) = 0$，其中 $E(\cdot)$ 是期望算子，那么

$$E(\mathbf{y}) = (\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\alpha} + (\mathbf{I} - \mathbf{B})^{-1}E(\mathbf{x})$$

$$= (\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\alpha} + (\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\mu}\mathbf{x} \quad (9)$$

和

$$E(\mathbf{y},\mathbf{x}) = \sum_{yx} = \begin{bmatrix} E(\mathbf{yy}') & E(\mathbf{yx}') \\ E(\mathbf{x}'\mathbf{y}) & E(\mathbf{x}'\mathbf{x}) \end{bmatrix}$$

$$= \begin{bmatrix} (\mathbf{I} - \mathbf{B})^{-1}(\boldsymbol{\Gamma\Phi\Gamma}' + \boldsymbol{\Psi})(\mathbf{I} - \mathbf{B})'^{-1} & (\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\Gamma\Phi} \\ \boldsymbol{\Phi\Gamma}'(\mathbf{I} - \mathbf{B})'^{-1} & \boldsymbol{\Phi} \end{bmatrix} \quad (10)$$

因此,方程(9)和方程(10)显示的结构方程模型,代表了一种在均值向量和协方差矩阵上的结构。这种结构是以模型的参数为根据的。协方差结构模型涉及方程(10)的元素。

从上面的讨论可知,协方差结构代表的统计模型有着广泛的通用性。协方差结构表示法给我们打开了解决模型识别问题的方便之门。尽管在本百科全书的很多词条都谈及了模型的识别问题,但是模型参数识别的一个必要条件是模型参数的数目要小于或等于协方差矩阵的元素数。特别是要对向量 $\boldsymbol{\Theta}$ 中的未知的模型参数进行排列。接下来,我们来看一个总体协方差矩阵 $\sum$,它的元素是总体方差和协方差。它假定一个潜在的实质性模型被设定用来解释在 $\sum$ 中的方差和协方差。我们知道在 $\sum$ 中的方差和协方差可用它们在样本协方差矩阵 $\mathbf{S}$ 中的对应样本值,用简明的计算样本方差和协方差的公式来估计的。因此,在 $\sum$ 中的参数是已识别的。

在确定 $\sum$ 中的元素已从对应的样本值被识别之后,我们需要进一步确定模型的参数是否已识别。如果它们可用协方差矩阵 $\sum$ 的元素被唯一地表示的话,那么我们说在 $\boldsymbol{\Omega}$ 中的元素就是已识别的。如果 $\boldsymbol{\Theta}$ 中的所有元素是已识别的,那么我们说模型就也是已识别的。

——David W. Kaplan

(沈崇麟译校)

## 参考文献

Browne, M. W. (1982). Covariance structures. In D. M. Hawkins (Ed.), *Topics in applied multivariate analysis* (pp. 72-141). Cambridge, UK: Cambridge University Press.

# 协变量(Covariate)

在统计分析中,协变量一词一般涉及的都是自变量【Independent Variable】。例如,在实验【Experiment】分析中,协变量是一个虽然不受实验者操控,但却对因变量仍然有着影响的变量。在存活率【Survival】或事件史分析【Event History Analysis】中,协变量无非就是一个连续的或非连续的,可能会对风险率【Hazard Rate】有影响的自变量,如一个时变(或时依)性协变量。然而,在对数线性建模【Log-Linear Modeling】中,我们却是通过将一种数值度量赋予定类变量,将连续变量作为"格协变量"处理,而不是使用真的连续变量。

——Tim Futing Liao

(沈崇麟译校)

## 表面故事(Cover Story)

在一些实验室研究中,调查的特定假设是比较明显的。如果参与者倾向于协助研究者确认已经显而易见的假设,那么这种研究目的的透明性就可能有问题。例如,假设研究者要求参与者完成一份问卷,其中包含了一系列问题,是关于他们对乌兹别克斯坦人的看法。在做问卷之前,研究者给参与者观看了一段有关英勇的乌兹别克斯坦人反抗恐怖分子,保卫家园的录像。然后再向参与者呈现这份态度问卷。在这种情况下,调查的目的似乎是显而易见的。有关需求特征【Demand Characteristics】的研究表明,过于明显的研究目标可能导致不真实的结果。一些参与者可能认为该研究的目的是为了引导他们对乌兹别克斯坦人作出正面的评价,即使那段录像(自变量【Independent Variable】)对他们感受的真实影响很小,他们也有可能会主动配合。

为了抵消由过度配合的参与者带来扭曲结果的可能性,心理学家会或多或少地精心设计一些表面故事来隐藏研究的真实目的。在这个例子中,研究者可能会略过预调查这一环节,告知参与者他们的任务是评价即将看到的录像的技术质量。在完成这个任务后,研究者可能会对参与者进行一个后测试,以了解他们对乌兹别克斯坦人的感情。作为表面故事的补充,这个任务的呈现方式可能也会隐藏研究的真实性质。接着,研究者会将这些参与者的回答与只进行后测试/预设对照组(posttest-only/comparison group design)的参与者的回答进行对比,那些对照组的参与者可能没有看过那段录像,或者看到的是一段不同的录像(Crano & Brewer, 2002)。

表面故事隐藏了研究的真实目的,因而是具有欺骗性的。因此,研究者必须仔细考量对它的应用是否符合研究伦理要求。我们为表面故事辩护的原因在于,它本质上是为了以对参与者最小的代价来确保实验结果的有效性,而非为了欺骗。一般来讲,表面故事多用于实验室实验【Laboratory Experiment】,在此类研究中,由于参与者的过度合作而造成令人误解的研究发现的危险最为严重。

科拉诺和布鲁尔(Crano & Brewer, 2002)区分了研究参与者的三种一般类型:自愿的、不自愿的和非自愿的。非自愿的参与者指的是他们没有意识到自己是科学调查的对象,所以在使用这样的参与者进行调查时,没有必要使用表面故事。自愿的参与者和不自愿的参与者都对自身参与调查的状态有所了解。自愿的参与者非常满意自己能够为研究贡献力量,如果他们试图帮助研究者确认研究假设,他们的积极态度就会带来对结论有效性的危害。不自愿的参与者也了解自己参与调查的角色,但他们并不为此感到高兴,常常认为自己是被迫参与调查。这些参与者的消极态度可能转化为试图毁掉研究的行为。在这种情况中,由于表面故事能掩盖研究的真实意图,故而对研究也有一定的用处。

——William D. Crano

(郑晓娟译　沈崇麟校)

### 参考文献

Crano, W. D., & Brewer, M. B. (2002). *Principles and methods of social research* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum.

# 隐蔽研究(Covert Research)

隐蔽研究意味着不告知被研究者或他们的合法代理人,他们正在被研究的基本信息。所有与社会研究相关的专业组织都会在自己的章程中包含知情同意【Informed Consent】的原则,并且他们一致建议反对隐蔽研究方法。

## 伦理问题

社会研究者应抵制隐蔽研究的原因是非常有说服力的。

首先也是最重要的原因即自主性问题。被研究者拥有知情、控制自己被观测的行为和根据参与调查的所冒风险或所得收益作出自己决定的权利。不仅如此,他们提供的资料是有价值的,因而他们有权利为研究者从他们那取得的资料进行定价,因此,如果研究者不告知被研究者他们所提供的资料情况,可以被视为对他们的掠夺或盗窃。

另外,使用非公开的研究方法可能使研究者获得坏名声。曾经发生过一个案例,亨利和哈布尔(Henle & Hubble,1938)偷偷藏在学生宿舍的床底下对学生对话进行录音,后来,劳德·汉弗莱斯(Humphreys,1975)曾装作看厕所的人来观察公共卫生间中同性恋者的邂逅。但是在今天,那些研究人类行为的学者则对偷听者或窥探隐私者这样的名声深恶痛绝。社会科学和行为科学专业组织表明了自己的道德地位,并把隐蔽研究丢弃给了做调查的新闻记者。他们也认识到,尽管行为被观测的范围可能是公开的,但行为本身则可能是私密的。

## 辩解

尽管专业组织和章程在隐蔽研究问题上形成了共识,但关于隐蔽研究方法是否可以接受这一问题仍然存在争论(Bulmer,1982; Homan,1992)。要让每个被研究者都明确获知他们正在被观察,并给予他们拒绝的选择权是不太现实的。社会科学中一些最有洞察力的研究作品,采取的都是对日常生活进行反思性研究的形式。欧文·戈夫曼将在沙滩和火车上放一块毛巾,或一张报纸的行为解释为宣示自己领地的行为:不会有人去寻求被观察者的同意或者贴出一张告示写着:“注意:现在正有一个社会学家在观察这个沙滩。”一个学习人类行为的学生一旦获得了赖特·米尔斯(Mills,1959)所说的“社会学的想象力”,他就会成为习惯性的研究者:研究行为不再是资料的收集,而是对资料的持续建构过程,而反对这种做法是无方法可循的。

获知研究正在进行并因而选择拒绝或避免的权利本身也是有问题的。例如,我们能够对于活着的人赋予这样的权利,而对于那些出现在区县档案局文档中已去世的人,这样的权利已无从谈起。拒绝研究的权利也有可能与公众利益背道而驰。例如,有一种颇有争议的观点认为,社会研究者应该对记者和警察的调查产生补充作用,去揭露儿童和弱势群体遭受虐待,或种族主义攻击得到资助的情况。的确,当社会研究学者准备接受被研究者说“不”的时候,他们的学术自由和学术品格对此作出了妥协。令人感到遗憾的是,处于权力地位的被调查者比那些相对无权、相对弱势的群体更倾向于使用知情同意权,把公开调查者拒之门外;因此,公开性伦理存在对弱势群体的歧视。

在某些情况下,隐蔽研究法很可能会得

不到任何反应。在这两种情况中,事实的确如此。第一,由于研究参与者不知道他们正在被记录,他们就没有理由为了获得一份赞许报告而调整自己的行为或数据。第二,研究行为不会抑制或贬低被调查者的话语。例如,在一个私密的祈祷会上,被调查者需要有自然的状态,如果知道了他们当中有人仅仅是来进行研究观察的,反而会对他们造成困扰。克服被告知研究的参与者给予反馈的这类问题的一种方法是,研究者与被研究者相处足够长的时间,使大家接受研究者成为研究场景的自然组成部分;另一种常用策略是,利用看门人和了解情况的合作者,他们并非是被研究者群体中的成员,但却能够为他们发言。

## 评论

我们很难在隐蔽研究和公开研究方法之间划分明确界线。即使使用最明晰的研究方法,例如,用访谈和问卷进行调查的研究者,也会采用一些重新定义研究行为的技巧来减少被研究者对研究的觉察。他们可能会询问被调查者一些关于他人的问题,而其他人对调查并不知情同意,在这种情况下,被调查者成为被代理的研究参与者,而他们本人并没有真的有效参与到研究中。尽管在与专业守则一致这一意义上,公开调查可以说是合乎伦理的,但公开研究方法并不必然在道德上更加高尚。

比关于隐蔽研究法的争论更为重要的一个问题是专业性的一致同意。知情同意是关乎社会研究的自尊与声誉的基本准则,从业者若公然藐视它,其后果将会十分严重。

十几年前,对隐蔽研究的标准反对理由在于隐蔽研究是对个人自由和相互信任,这些最为重要的价值理念的损害和背叛。这些年我们看到了公众场所隐蔽摄像头的出现,看到了具备技术能力的人对计算机记录和数据库的入侵,看到了高层人士的欺骗和相互攻击。信任和自由的环境已经发生了改变。社会研究者努力发现的结论可以被政客扭曲利用。因此,研究者的可信度是否站得住脚,可能已不再取决于他们所使用研究方法的公开程度,而是取决于他们能够避免控制和操纵,进而提供使用者信任的结论的程度。

——Roger Homan

(郑晓娟译　沈崇麟校)

## 参考文献

Bulmer, M. (Ed.). (1982). *Social research ethics.* London: Macmillan.

Henle, M., & Hubble, M. B. (1938). Egocentricity in adult conversation. *Journal of Social Psychology*, *9*, 227-234.

Homan, R. (1992). *The ethics of social research.* London: Longman.

Humphreys, L. (1975). *Tearoom trade: Impersonal sex in public places.* Chicago: Aldine.

Mills, C. W. (1959). *The sociological imagination.* New York: Oxford University Press.

# Cramér *V*

参见对称量度【Symmetric Measures】。

## 创作式分析性实践(CAP)民族志(Creative Analytical Practice [CAP] Ethnography)

自 1980 年代以来,在后现代主义【Postmodernism】、女性主义以及后结构主义【Poststructuralism】的影响之下,当代的民族志学者【Ethnographers】对标准的社会科学写作发出了挑战,认为它在伦理、哲学和科学上都有问题。这些当代民族志学者承认在"事实"和"虚构"之间、"主观"和"客观"之间、"作家"和"主题"之间的界限是模糊的,他们要用新的方式来进行社会科学写作。这个工作最初被称为实验民族志或另类民族志。最近,为了强调完成这种写作的过程,并给这些民族志建立其自身的合法性,它们更多地被称为创作式分析性实践民族志或 CAP 民族志。

CAP 民族志既是"科学的"——(就忠实于以实证工作和研究了解的这个世界而言),又是"文学的"——(就用引人情感共鸣的写作技巧和形式来表达其所知而言)。尽管 CAP 民族志可归属于许多不同的学科,它们在研究/写作实践方面却有一些共通性,包括运用理论来挑战学科权威的根基,写对社会和个人重要的主题,重视无等级性的、合作的和/或多重意义的写作,喜欢自反性,把自己置于多重的甚至是竞争性的话语中,还有签名的做法,引人情感共鸣的写作,以及用不同形式为不同的读者写作。

CAP 民族志的种类很多(参见 Richardson, 1997),包括自传式民族志【Autoethnography】(Ellis, 2003)、小说【Fiction】、诗歌式的表现形式、表演文本、民族志戏剧、故事写作、读者剧场、格言、谈话、书信集、多媒体、分层叙述,以及超文本、博物馆展览、照片、绘画、舞蹈及其他艺术表现形式。例如,劳雷尔·理查森曾根据对一位未婚妈妈"Louisa May"的开放式采访创作了叙事诗(Richardson, 1997)。卡罗琳·埃利斯(Ellis, 2003)发明了一个虚拟的研究生研讨会,会上以现实中的学生为基础综合而成的各个角色进行了一场关于自传式民族志的讨论。

### 对 CAP 民族志的评价

大多数 CAP 民族志都遵循六大标准:

1. 实质性:这篇东西是否对我们理解社会生活有实质性的贡献?

2. 审美性:这篇东西是否会招致诠释反应?它在艺术上是否令人满意,是否复杂?

3. 反身性(reflexivity):作为该文本的生产者,作者是否反思自我?

4. 伦理标准(ethics):作者是否认为自己应该为了解和讲述其研究对象负责任?

5. 影响:该文本是否会给读者带来影响?在情绪上影响?在认知上影响?是否会让其他人去开展社会行动?这个文本是否会对作者有所改变?

6. 表达现实:该文本是否提供了一个文化、社会、个人或社区意义上的"现实"的可靠的解释?

CAP 民族志学者承认作品不能和作者分离——主体性、权威性、作者身份、反思性等错综复杂地纠缠在一起,编织成了这种表达形式。科学是一种视角,创造性的艺术是另一种视角。CAP 民族志学者尝试通过两种视角一起来看。

——Laurel Richardson

(杨可译 沈崇麟校)

## 参考文献

Denzin, N., & Lincoln, Y. (Eds.). (2000). *Handbook of qualitative research* (2nd ed.). Thousand Oaks, CA: Sage.

Ellis, C. (2003). *The ethnographic "I": A methodological novel on doing autoethnography.* Walnut Creek, CA: AltaMira.

Richardson, L. (1997). *Fields of play: Constructing an academic life.* New Brunswick, NJ: Rutgers University Press.

# Cressie-Read 统计量(Cressie-Read Statistic)

Cressie-Read 统计量属于效力发散统计量族(family of power divergence statistics),用以评估模型的拟合,它是似然比统计量【Likelihood Ratio Statistic】的一个备择统计量,但它没有一个与之关联的 $p$ 值。

——Tim Futing Liao

(沈崇麟译校)

## 参考文献

Cressie, N., & Read, T. (1984). Multinomial goodness of tests. *Journal of Royal Statistical Society Series B*, *46*, 440-464.

# 效标效度(Criterion Related Validity)

在教育评估和社会科学测量中,与标准相关的效度是在测试或测量验证的背景下定义的。测试和测量验证是为了确认根据测试分数或测量结果做的推论的可信度而进行的收集数据过程——一种关于分数和度量的推论(如同决策)与一种关于分数和度量的解释(意义)的推论。在 1985 年以前,《教育与心理测验标准》(*Standards for Educational and Psychological Testing*)(American Educational Research Association, 1985)效度是以行为主义者的传统概念化的,在分类学上被分成了彼此有着较大区别的三大类,即内容效度【Content Validity】、效标效度【Criterion Related Validity】和构念效度【Construct Validity】。这样三类效度法为我们提供了一种测试或量度工具的累积的证据。效标效度是一种测试或工具与某种效标之间关联的证据。这种效度(检测)法普遍用于才能测试(aptitude tests)或其他一些场合,在这些场合中,测试分数或量度被用来做预测。效标效度还可进一步被分为两种子类:(a)效标效度预测证据(或预测效度);(b)效标效度同发证据(或同发效度)。效标效度的预测证据表明一种测试或工具在预测一个应试者今后进行某一特定的活动(能力)的有效性。例如,如果要声称某种竖琴(演奏)能力测试具有预测效度,那么测试方法的开发者就必须要证明在这项测试的高分者具有高水平的竖琴演奏能力;反之,低得分者竖琴演奏水平相对较低。

效标效度的同发证据表明测试与某种效标关联的程度。同发效度与预测效度唯一不同之处就是这个效标是同时的。例如,

某一设计用于诊断心理压抑的测试，在将它实施于一个我们已经知道他（她）的罹患抑郁症程度的患者时，可能是有效的。同发效度常用来替代预测效度。因为它不会延误效标的实用性。现在，我们来举一个例子。如果一个量表被构建来评估学生的科学能力，那么将通过测试分数和学生目前自然科学课程的成绩联系起来，而不是与将来的科学能力的指标（如科学技能、科学论文的写作等）联系起来，可能就可以建立这种同发效度。这两种效度之间的差别在于测试的目的。如果一个测试或工具要用于预测一个效标的品质或性状，那么就需要用预测效度。如果一个测试或工具被用来确认一个效标的品质或性状，同发效度则更佳。关于效度概念化最新的一些思考，都反映在了1985年和1999年的《教育与心理测验标准》中了。它强调有效（或无效）的不是测试，而是有关意义的推论和测试分数的使用。“效度涉及的是证据或理论对使用测试提出的测试分蕴含的解释的支持程度”（American Educational Research Association, 1999, p. 9）。有人认为效度是一个“统一的，但只是某一方面的概念”（Messick, 1989）。对它而言，不同类型的证据（不是不同类型的效度）应该聚合在一起，以支持基于测试分数（或量度）作出的推论和测试（或工具）的恰当使用。上面讨论的效标效度的事例，都是这种类型的证据的例子。

——Chao-Ying Joanne Peng
Daniel J. Mueller
（沈崇麟译校）

＊也可参见构念效度【Construct Validity】、效度【Validity】。

## 参考文献


American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1985). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.

Anastasi, A. (1988). *Psychological testing* (6th ed.). New York: Macmillan.

Kane, M. (2002). Validating high-stakes testing programs. *Educational Measurement: Issues and Practice*, *21*(1), 31-41.

Messick, S. (1989, March). Meaning and values in test validation: The science and ethics of assessment. *Educational Researcher*, pp.5-11.

Messick, S. (1992). Validity of test interpretation and use. In M. C. Alkin (Ed.), *Encyclopedia of educational research* (6th ed., pp. 1487-1495). New York: Macmillan.


# 批判话语分析（Critical Discourse Analysis）

批判话语分析（CDA）把语言当成社会事件、社会实践的一个要素，辩证地与其他要素（包括社会制度和物质世界的各个方面）相关联。其目标是说明语言与其他的社会事件、实践之间的关系。这是一个源于语言学的学习语言分析的方法，但比大多数语言学都更加面向社会，也更跨学科。

在这里，话语的意思是把语言（以及“身体语言”、视觉图像等）看成社会的一个瞬间（moment）。这种方法之所以是“批判

的”,是就以下意义而言:(a)它试图说明语言和社会生活其他要素之间不明确的关联;(b)它集中关注的是语言如何在权力、控制、剥削的社会关系构建和复制中起作用;和(c)它对社会解放与推进社会公正问题予以特别的语言关注。在这些方面,批判话语分析可以被视为批判社会科学的一个分支。

CDA 基于一个假设:社会事件的语言要素(谈话、文本)可能会促成其他社会要素发生改变——话语是社会建构的。分析的一个焦点就是话语对于意识形态的建构、复制和改变所起的作用。这两点都符合一个辩证的观点:话语是一种社会要素,它与其他社会要素不同,但却不是离散的——不同的要素彼此“内化”。可以说,社会生活近来的变化之一就是话语在与其他要素相关的某些方面变得更为突出了;例如,“知识”或“信息”社会的概念似乎暗含着一个意思,即社会(如组织)变革可能是由可以创生新的行为方式、灌输新的认同,在新的工厂里具体实现生产的话语“引导”的。不过,比起某种话语唯心主义,我们应该采取社会建构论【Constructivism】的温和观点,承认社会构念具有无法传递的特性,可能会使它们抗拒新的话语的内化。

## 该领域的发展

对于语言与社会生活其他要素之关系的批判观点,可以追溯到亚里士多德的修辞学【Rhetoric】,同时也见于许多学科中。尽管 CDA 处理诸如这样的议题并非是独创的,但它还是首次开发出了一个相对系统的理论和研究的主体。它最主要的资源包括语言学中广义的“功能主义”(与“形式主义”相对)传统,包括米歇尔·哈利迪的系统功能语言学理论;西方马克思主义,尤其是葛兰西和阿尔都塞有关“霸权”和“意识形态”的理论;法兰克福学派的“批判理论”,包括最近的哈贝马斯;福柯的“话语”概念;以及巴赫金的对话语言理论,包括其对“异质性(heteroglossia)”和“互文性(intertextuality)”的强调与“类型(genre)”的理论化。

CDA 这个术语 1985 年前后首次使用,但 CDA 可以被视为包括了各种方法,首先是批判语言学,这是对哈利迪从意识形态和权力角度来分析文本的语言学的一种应用。有些批判语言学的作品(尤其是特恩·迪杰克)的作品结合了认知心理学,有些着重于历史文献(尤其是鲁思·沃达克的“话语—历史”方法),有些(尤其是甘特·克莱斯和西奥·范·莱文)集中关注当代文本的“多模式(multimodal)”特点,尤其是在语言和视觉图像混用时。CDA 越来越跨学科,加入了更多的社会理论和研究,而且(尤其针对费尔克劳而言)致力于提高其对当代世界社会转型(全球化、新自由主义、新资本主义)的研究能力,以表明语言是如何在社会转型过程中发挥作用的。

## 应用

CDA 近来在社会科学的许多学科和各种领域的研究中得到采用(例如,地理学与城市研究、健康科学、媒体研究、教育研究、社会学)。在这些广泛的应用中,主要可分为两种研究:一种是结合了精细的语言学文本分析的研究;另一种是基本按照福柯的话语观、利用话语分析的理论视角而非那种精确的文本分析来做研究。对有的 CDA 研究者来说,在社会研究中确立精细的文本分析的价值是其中心关注点。与此同时,他们也考虑如何从公共领域内的社会和政治辩论中找到语言的焦点(Fairclough,2000)。

## 例子

有一个主题已经吸引了全世界许多关注的目光,现在已经有了自己的虚拟研究网络,这就是新资本主义的语言(language in new capitalism),它探讨的是在如今对资本主义的重组和"全球性的"重新调节中语言是如何发挥作用的。对全球经济变迁叙事的分析可以作为例子来说明 CDA 可以促进对这些转变的社会研究(Fairclough,2000)。这些叙事在经济、政治及其他类型的文本中非常普遍。这些文本中循环往复地表现经济变迁,将其当作一个不需要人类主体参与的过程——因此,打个比方,新的市场"开放了",而不是人们开放了市场,而且技术也从这儿"迁移"到那儿,而不是由公司带着去的——这个过程在时间和空间上都是普遍性的(它平均地发生在每个地方、每个普遍意义的当下)。这种对经济变迁的表达常常与政策上的禁令一起出现——无法论证、不可避免的实际情况"是"什么,决定了我们"必须"在政治的意义上去做什么。对这些叙事的分析可以为布迪厄提出的框架性的讨论上增添一些血肉。布迪厄曾提出,新自由主义的话语是开创"全球化"(以及资本主义转型)的军械库中一个有力的组成部分,而认为全球化已然完全实现其实是一种误导。

——Norman Fairclough

(杨可译 沈崇麟校)

## 参考文献


Chouliaraki, L., & Fairclough, N. (1999). *Discourse in late modernity*. Edinburgh, UK: Edinburgh University Press.

Fairclough, N. (2000). *New labour, new language*? London:Routledge Kegan Paul.

Fairclough, N., & Wodak, R. (1997). Critical discourse analysis. In T. van Dijk (Ed.), *Discourse as social interaction*.Thousand Oaks, CA: Sage.

Wodak, R., & Meyer, M. (2001). *Methods in critical discourse analysis*. Thousand Oaks, CA: Sage.


# 批判民族志(Critical Ethnography)

近年来,批判民族志已日益变成了一种流行的视角。

不过,不同学科的学者使用不同的理论方法和范式对它进行的任意剪裁为其带来了广泛的应用,但这导致任何形式的文化批判都被贴上了批判民族志的标签。因此,确定一些基本的组成要素,可能比起定义批判民族志更为重要。

将批判民族志简单地还原成社会批判扭曲和简化了批判民族志的事业。从最简单的意义来说,批判民族志是一种将颠覆性的世界观用于更为传统的文化研究叙事的方式。它不一定与传统的民族志【Ethnography】或传统的社会科学针锋相对。相反,它倒是提供了一种对知识与社会的关系的较具反思风格的思考,并免除了不必要的社会统治形式。

批判民族志不同于其他种类的民族志,并不是批评(criticism),而是评论(critique)。批评是我们发现蛋没有煮熟时发出的抱怨,一般来说,它意味着对一种已知事态的不满,但不一定带着对(事态)表面之下的东西进行深挖,以确定基本的社会过程加以挑战的责任。

与之相反,评论则是用一个额外的假设,即如果我们对潜在的条件来源,包括我们的抱怨所依据的价值观进行考察,那么"事情可能会更好"这样一个假设来评估

“事情究竟如何”。

批判民族志不是一种理论,因为它自身并没有生成一套可以检验的命题或法则式的陈述。相反,它是一种提供了有关我们社会世界的基本图像、比喻和理解的观点。它还提供了一些价值前提,我们可围绕这些前提提出问题,对如何改变提出建议。它是一种“努力撕下霸权的假面具和揭露压迫势力”的方法论(Crotty,1998,p.12)。批判的研究者都从所有的文化生活都处于控制与抵抗的恒定张力中这个前提出发。这种张力反映在行为、互动仪式、规范体系和社会结构之中,在组成既定文化的规则、交流系统和人工产品中都可以看到。批判民族志关注那些看似日常的,甚至令人厌恶的事件,并将其再现出来,展现出广义的控制社会过程、权力不平衡和那些将一套偏爱的意义或行为强加于他人的符号机制。尽管并非所有的批判民族志都欢迎概推性【Generalizability】、信度【Reliability】、效度【Validity】、确证【Verification】等问题,但批判民族志还是与“经验科学”的规范有颇多相似之处。依靠这些规范,经验之谈的信度才可以被评估(Maines,2001)。

有两个例子可以说明批判民族志的多样性。第一个是博斯沃思(Bosworth,1999)对三个英国女子监狱文化的批判民族志研究,该研究发现了一个悖论:传统上曾作为惩罚过程一部分的理想化的异性恋女性的形象,也成了女性反对控制并熬过牢狱生活的一个手段。她提出,尽管女子监狱中很多日常事务都基于对女性气质的理想化观念,但它也有一种矛盾和讽刺的效果,即制造抵抗的可能性(Bosworth,1999,p.107)。这种评论颠覆了长期以来通过男性囚犯和男性学者的棱镜看女子监狱文化的传统,减轻了刑事司法研究中参差不齐的性别差异。

费雷尔和哈姆(Ferrell & Hamm,1998)收集的颇具洞见的例子可以进一步说明批判民族志的用途。通过研究诸如皮条客、大麻吸食者、色情电话服务员、国内的恐怖分子等边缘人群,作者鼓励读者去改造有关犯罪、越轨和社会控制的惯有的、舒服的假设和形象。这反过来又鼓励我们去更充分地考察创造和维持社会文化权力不平衡的不对称的权力关系。

当然,批判民族志的确存在一些问题(Thomas,1992)。但科学——而且评论是科学事业的一部分——是会自我校正的。从事民族志评论的学者其任务之一就是要在先验的概念分析、阐释性的叙事、实证主义【Positivism】的精确、理论建构等这些有时候充满矛盾的任务之间保持平衡,小心不要像实证主义者那样拒绝这些任务,以免我们用一种知识教条来代替另一种。这也意味着我们必须不断重新评价我们的事业,就像评价我们的分析焦点那样严格,而且一直牢记事物从来不是表面上看来的那样,而且社会公正并不只是一个目标而已,而是一种天职,所有的学者都应该用可证实的评论来为之奋斗。

——Jim Thomas

(杨可译　沈崇麟校)

## 参考文献

Bosworth, M. (1999). *Engendering resistance: Agency and power in women's prisons.* Aldershot, UK: Ashgate/Dartmouth.

Crotty, M. (1998). *The foundations of social research: Meaning and perspective in the research process.* Thousand Oaks, CA: Sage.

Ferrell, J., & Hamm, M. S. (1998). True confessions: Crime, deviance, and field research. In J. Ferrell & M. Hamm (Eds.), *Ethnography at the edge: Crime, deviance and field research* (pp. 2-19). Boston: Northeastern University Press.

Kirk, J., & Miller, M. L. (1986). *Reliability and validity in qualitative research.* Beverly Hills, CA:

Sage.
Maines, D. R. (2001). *The faultline of consciousness: A view of interactionism in sociology*. New York: Aldine De Gruyter.
Thomas, J. (1992). *Doing critical ethnography*. Newbury Park, CA: Sage.

## 批判诠释学(Critical Hermeneutics)

批判诠释学是在保罗·利科(Ricoeur, 1981;Thompson,1981)的诠释现象学以及约翰·汤普森(Thompson,1990)的深度诠释学(depth hermeneutics)的基础上发展而来的。沿着利科的脉络,批判诠释学把解释分解为一个辩证的过程,涉及三个相互关联又相互独立、互相修正的阶段或是“环节”,即社会历史分析的环节、形式分析的环节及解释—再解释的环节。前面两个分析环节都汇聚在这里(实证的例子可参见 Phillips & Brown,1993)。在解释一个文本时,我们可以且应该分析它所在的社会历史环境。但我们也应该从形式上分析这个文本,用一种文本分析的形式方法,例如,结构符号学【Semiotics】或者话语分析【Discourse Analysis】,将其当作从经验中抽取出来的符号系统来分析。从批判诠释学的视角来看,分析社会历史环境或文本本身都不完美。这两者都是解释过程中不可或缺的环节,那些未能明确包含社会历史环境的解释方法和分析文本形式的解释方法都是不完整的。研究者在完成了社会历史和形式分析之后,必须创造性地把这两个环节结合起来,对文本以及它在其所属的社会系统中的角色给出一个解释。在这个解释环节中,研究者才对这个文本在持续不断的社会实在的建构中扮演什么角色发展出深刻的理解。这个分解为三个环节的过程给读者提供了一个结构化的框架,让读者可以沿着这个框架更明确地理解研究者的方法论,并对最终结论的价值作出清楚的判断。

但是这种方法论除了阐释以外,也是批判的。除了关注意义之外,批判诠释学还关注一个问题,即在这个社会世界的典型特征——不对称权力关系的不断维持中,沟通的角色是什么?对文本产生、传递和接受的社会历史环境的考察,给研究者提供了一个关注沟通在导致机会和生活机遇长期不对称的权力关系发展中的角色的机会。批判诠释学引入了解放旨趣(emancipatory interest),把社会系统背后想当然的惯例和规则当作关注的焦点。使用批判诠释学的研究者探询特定社会环境中沟通的功用,沟通是为了谁的利益,以及什么人可以参与其中。

批判诠释学将解释文本的旨趣与解释社会文本的社会历史环境的旨趣相结合,作出了三个重大的贡献。首先,它提供了一种结构化的办法来检验社会生活中符号现象的作用。其次,它提供了一个结构化的办法来检验文本的材料来源以及其创作者的利益所在。最后,它提供了将文本分析的形式方法和对文本产生的社会环境的兴趣相结合的办法。因此,批判诠释学在社会科学较为传统的解释方法之上又增加了一个重要的维度。

——Nelson Phillips

(杨可译　沈崇麟校)

## 参考文献

Phillips, N., & Brown, J. (1993). Analyzing communication in and around organizations: A critical hermeneutic approach. *Academy of Management Journal*, *36*(6), 1547-1576.

Ricoeur, P. (1981). *Hermeneutics and the human sciences: Essays on language, action and interpretation.* New York: Cambridge University Press.

Thompson, J. (1981). *Critical hermeneutics: A study in the thought of Paul Ricoeur and Jurgen Habermas.* Cambridge, UK: Cambridge University Press.

Thompson, J. (1990). *Ideology and modern culture.* Stanford, CA: Stanford University Press.

# 关键事件法(Critical Incident Technique)

关键事件法(CIT)大约半个世纪以前被设计、应用于一项科学研究中(Flanagan, 1954)。这种方法采取当时在社会科学以及自然科学中作为主导范式的实证主义的方式。它主要被用于职业环境中,其提倡者确立了它的效度【Validity】和信度【Reliability】(Andersson & Nilsson, 1964; Ronan & Latham, 1974)。弗拉纳根的方法以及他所评述的研究都假设现实是具体实在的。与此相反,切尔(Chell, 1998)所发展的CIT却采用了现象学【Phenomenology】。其目标是从技术上通过非结构化访谈【Unstructured Interview】的办法来获取对被访者有意义的事件的思想过程、参照体系和感觉。访谈要求被调查者【Respondent】结合他们的生活环境、当前的情况、态度和取向对事件的意义进行陈情【Account】。这种方法可以被用于深入的个案研究【Case Study】以及多调查点的调查之中。

CIT是一个定性访问【Qualitative Interview】程序,它有助于对被访人所认定的重要境遇(事件、事变、过程、议题)进行研究,从认知效果的角度考察这些事件是如何形成的、结果如何。研究目标是从个人的视角获得对事件的理解,同时兼顾到认知、情感和行为要素(Chell, 1998, p.56)。

## 例子与方法

可以说,所有的研究几乎都是从“初步设计工作”开始的。这假定了研究自有其价值;研究问题和想法都是前后一致的,并聚焦于研究目标。不仅如此,在经验工作中,研究者需要考虑如何找到被访人(或被试者),以及为什么这位特定的被访人会被选中。研究者可利用他们的地方性知识、口碑、其他人的数据库和列表等各种资源。这也意味着可能会有一个选拔被试者的标准,并且需要对此进行明确的说明。对商界小型企业的个人问题和生意问题的研究可以作为使用CIT有助于展示信息丰富程度的一个例子。这个例子说明了个人和家庭问题相对于其他的生意和经济因素的重要性(Chell, 1998; Chell & Baines, 1998)。

与确定被访人紧密相关的一个问题是获取调查通道【Access】的能力。这可能需要技巧、耐心和毅力。不要把接近他们当成想当然的事。一开始,可以使用打一个自我介绍的电话或者写一封介绍信的办法。后面这个办法可以让研究者简要地介绍项目的宗旨和目标,为占用研究被调查者的时间和介入他们的隐私【Privacy】提供可信的理由。所以说,从一开始我们就需要谨慎地处

理问题。

一旦找到了进入通道,并与被访人约定,研究者就需要“排练”CIT 的访谈。尽管访谈是非结构化的,但还是有一定的顺序,首先访问员要介绍他/她自己及其研究,并向被访人介绍 CIT 的特点。保密性【Confidentiality】以及相关的保证是赢得被访人信任的根本。一旦被访人放松下来,并不再对研究程序感到不舒服了就可以开始访谈。

接下来就会要求被访人集中谈一个或更多个(通常不多于四个)发生在他自己身上或自己十分熟悉的群体、组织中的“事件”(在一个特定的时期)。可以用一些不同的策略来让被访人集中讲述关键事件。不同的被访人表述一个事件的能力差异很大,最难的访谈大概是那些似乎根本就没有事件的访谈!那么访问者就必须试着在看似缺乏事件的啰唆片段中去探究,生成有用的信息。下一步的困难是,确认正确的时间顺序或者事件的顺序。这可以借助其他事件、文献、日记以及其他有效的材料来加以推论。可以鼓励被访人提供这种交叉引用的材料。这会有助于信息的三角检验【Triangulation】和验证【Validation】。

被访人也可能会认为事件是正面的经验或负面的经验。被访人如何表达这个事件完全由他们掌握,事实确实如此,他们正是用语言来表达他们经历的事件的。尽管研究者捕捉可评估的语言的确很重要,但通过仔细地倾听、适当地追问【Probes】,从被访人的角度抓住事件的基本细节同样也是很重要的。

尽管非结构化访谈是结构性的,但对访谈要有所控制还是很重要的。访谈可以采用一般性的探查方式(例如,接下来发生了什么?你怎么处理的?),这会引出有关环境、语言和关系的信息。如果访问员不理解,就必须插话,立即得到说明;否则这条线索可能就断了,通常访谈结束时要回到一些小的点就很困难了。要记住那些没有理解的东西可能也是很困难的!最后,在结束 CIT 使用时,访问员必须考虑到所有的伦理问题。如果匿名【Anonymity】很重要,那么必须作出保证。如果某种方式的反馈很重要,那么必须保证做好反馈的安排。访问员必须在访谈结束离开的时候感觉到,无论将来什么时候、什么阶段回来,都会受到被访者的欢迎。

这个过程的最后步骤就是生成录音转录【Transcription】,然后再对它进行分析。分析过程很可能是以扎根理论【Grounded Theory】为基础。这意味着研究者要抛却先入之见,通过分析的过程,依靠对数据的概念化【Conceptualization】建立起解释框架。这个理论【Theory】建构的过程是其最基本的形式。理想情况下,它可以和理论抽样【Theoretical Sampling】技术一起运用,以便得到更深入的理解,并最终形成逻辑连贯的理论。CIT 运用得当并不容易,它要求研究者能熟练掌握这种方法。可能有些压力,但整体而言,对研究者和被访人来说都很愉快。这有助于帮我们理解社会交往中的动态过程。它认为现实生活是很混乱的,而所有的人,包括研究者和被访人都试图去理解他们的现实。

——Elizabeth Chell

(杨可译 沈崇麟校)

## 参考文献

Andersson, B. E., & Nilsson, S. G. (1964). Studies in the reliability and validity of the critical incident technique. *Journal of Applied Psychology*, *48*(1), 398-403.

Chell, E. (1998). Critical incident technique. In G. Symon & C. Cassell (Eds.), *Qualitative methods and analysis in organizational research: A practical guide* (pp. 51-72). London: Sage.

Chell, E., & Baines, S. (1998). Does gender affect

business performance? A study of micro-businesses in business services in the U. K. *International Journal of Entrepreneurship and Regional Development*, *10*(4), 117-135.

Flanagan, J. C. (1954). The critical incident technique. *Psychological Bulletin*, *51* (4), 327-358.

Ronan, W. W., & Latham, G. P. (1974). The reliability and validity of the critical incident technique: A closer look. *Studies in Personnel Psychology*, *6*(1), 33-64.

## 批判实用主义(Critical Pragmatism)

批判实用主义是一种方法论取向,它认为社会科学研究应当阐释意识形态的统治、霸权的行为和社会不公;在追求这样的阐释时,它还鼓吹一种折中的方法论实用主义(eclectic methodological experimentalism)。

批判实用主义的关键要素源自批判理论【Critical Theory】,特别是马尔库塞和葛兰西的理论。这两位作家都呼吁研究应关注社会自身如何再生产,以及人们是怎样被说服去接受那些明显不利于他们的观点、信仰和行为的。马尔库塞感兴趣的是人们如何学会挑战普遍存在的单向度的思维,也即那些禁止人们对社会的道德基础进行任何思考、劝阻人们不要进行乌托邦式幻想的思想(Marcuse, 1964)。葛兰西关注的是人们学会承认和质疑霸权的方式,也即人们接受拥抱对其有害、维护不平等结构的思想和行为的过程(Gramsci, 1971)。而实用主义的要素来自美国的实用主义【Pragmatism】传统。实用主义有一种实验主义的、反本质主义的味道,主张不拘一格采用各种方法和办法来发展民主。它提倡对假设进行经常的批判分析,而这就是对具体化、标准化的行为模式或探究方式保持怀疑。

在批判实用主义中的可误论【Fallibilistic】的自我批判的要素也意味着它避免依附于某种特定的方法论,无论是定量的、定性的、实证主义【Positivism】的、民族志【Ethnography】的、统计的,还是现象学【Phenomenology】的方法论。批判实用主义的立场拒绝过早陷入某种研究方法之中,无论它看起来多么宽松。它呼吁一种有原则的方法论的折中主义,为了解释和质疑主导性的意识形态愿意采用任何方法。

批判实用主义研究关注许多问题。人们是如何学会主导意识形态的各种推论方法的?他们如何学会质疑意识形态合理化了的社会、文化和政治形态?人们如何学会以突出自身和他人联系的方式阐释自身经验,并认识到团结和集体组织的必要性?人们如何撕下他们的生平和社区中的权力流动(the flow of power)的面具?人们是如何得悉霸权的存在,并了解他们自身在霸权持续存在中的共谋关系?一旦认识到了霸权,人们又是如何与其无所不在的权力效果相抗争的?约翰(Gaventa, 1980)对阿巴拉契亚(Appalachia)地权的研究可以作为一个定量取向的批判实用主义的例子。道格拉斯(Douglas, 1990)对得克萨斯州南部小镇的墨西哥裔美国青年的民族志研究可视为定性取向的例子。

——Stephen D. Brookfield

(杨可译　沈崇麟校)

### 参考文献

Foley, D. E. (1990). *Learning capitalist culture: Deep in the heart of Texas*. Philadelphia: University

of Pennsylvania Press.
Gaventa, J. (1980). *Power and powerlessness: Quiescence and rebellion in an Appalachian valley.* Chicago: University of Chicago Press.
Gramsci, A. (1971). *Selections from the prison notebooks.* London: Lawrence and Wishart.
Marcuse, H. (1964). *One dimensional man.* Boston: Beacon.

## 批判种族理论(Critical Race Theory)

批判种族理论是指对流行法律范式和现行法律中默认的白人至上的大量激烈批评。法律文献和有关公民权的话语曾经为白人学者和法律从业者的作品所垄断。批判种族理论自1990年代兴起,主要是(但并不仅仅是)由美国的有色人种进步学者所推动的思想运动(Crenshaw, Gotanda, Peller, & Thomas, 1995)。批判种族理论解构了代表白人的法律、象征和物质资本。它的使命就是批判法律,并最终对促成构建和维护种族统治和臣服关系的法律进行改造。尽管批判种族理论以法律学科为基础,但因为其综合了经验深度、政治以及理性的严谨,故而易于得到社会科学家的理解和重视。

批判种族理论可视为对抗性写作(oppositional writing)的激进传统的一部分(Essed & Goldberg, 2002)。这些方式预设了一个前提,即科学并非“中立”或“客观”的。对历史上曾经经受苦难的群体而言,权利和平等、对和错、公平和不公的概念有着不同的含义。方法论上有一个重要的问题,即对种族和种族主义常有不同的理解,这要取决于复杂的种族关系矩阵中,具有定义权一方的位置(Bulmer & Solomos, 2003)。包括法院在内的各种机构也倾向于原谅作恶者的种族主义观点,因为经常采用的是基于动机证明的一个狭义定义。而那些经受了各种种族主义之苦的受害人,通常都能理解排斥和羞辱的社会后果。

批判种族理论在将交叉性和多重认同作为理论框架主流化的过程中发挥着作用。种族、性别以及其他结构类型互相交叉,决定了法律和社会的话语。例如,在儿童福利体系中,非洲裔美国妇女失去孩子监管权的情况高得不成比例,其部分原因在于,文化偏见更倾向于白人中产阶级的核心家庭和(全职)母亲的理想。

批判种族理论的学者把文本和叙述作为数据,揭露出传统法律分析的限制和歪曲之处。他们通过对曾经作为种族关系标志的历史性案例以及其他法院的案例的语境分析解构了法律。著名的“隔离但平等”案,即普莱西诉弗格森案(Plessy *v.* Ferguson,1896)下面潜藏的色盲的迷思是一个很合适的例子。法庭把种族当作与社会地位无关的范畴来对待,重新确认了种族隔离制度。批判种族理论提供了一个理解法庭系统如何以及为何要继续强化种族不公的框架。对历史和当代有关种族不公的文本和故事进行收集和分析是非常有力的工具,有助于在追求社会变革的过程中进行确证和理解。

——Philomena Essed

(杨可译 沈崇麟校)

### 参考文献

Bulmer, M., & Solomos, J. (Eds.). (2003). *Researching race and racism.* London: Routledge Kegan Paul.
Crenshaw, K., Gotanda, N., Peller, G., & Thomas, K. (Eds.).(1995). *Critical race theory.* New York: The New York Press.

Essed, Ph., & Goldberg, T. D. (Eds.). (2002). *Race critical theories: Text and context.* Malden, MA: Blackwell.

Plessy v. Ferguson, 163 U.S. 537 (1896).

## 批判实在论(Critical Realism)

批判实在论是一种理解自然和人文社会科学本质的方式。尤其是对理解人类社会科学而言,批判实在论也提倡某些理解方式,并把它们应用于实际,以反对其他的一些理解方式。作为唯实论【Realism】的一支,它公开承认,科学知识的对象其存在和行动都独立于对它们的认识。不过,"批判的"这个限制条件,使得这种实在论区别于其他实在论,也即所谓的直接或朴素实在论。后者假定在信仰和现实之间有一种简单的"一一对应"关系。在人文社会科学中,限定词"批判的"通常还代表一种对既有的社会形态的一种规范化的批判取向。

批判实在论者在独立存在的实在、关系、过程等["不及物维度(intransitive dimension)"]科学知识的对象,与社会文化生产出来的概念、知识主张以及我们试图理解它们的方法之间作了清晰的区分。后者[有时被称为"及物的维度(transitive dimension)]"(Bhaskar, 1979/1998)总是临时性的、不可靠的。因此,批判实在论者充分承认社会科学理论框架和知识宣称由社会和文化"建构"的特性,但他们指出,诸如实验、科学教育以及实践科学观点等活动,以科学知识对象的独立存在为先决条件。除非将那些实践置于它们是在这一断言的基础之上,否则我们就无法理解这些实践,即使如此,我们也还是可以对有关"某物"的特定断言的真实性提出质疑。

大多数情况下,批判实在论者在反对相对主义者时,常与一些方法联系起来(例如他们自己采用的方法),这些相对主义强调知识的社会建构的特征。即使不对不可证实的"终极"或"绝对"真理的概念作出承诺,通常它也总是可以根据已有的证据,对一些竞争的知识宣称的理性正当程度(暂时的)进行评估。

批判实在论者对经验主义者和相对主义者对自然科学活动之本质的观点提出了强有力的挑战。特别是巴斯卡尔(Bhaskar, 1975/1997)用诸如科学实验这样的实践的先验论证(transcendental arguments)证明了在科学本体论【Ontology】和科学规律描述上的经验限制的不连贯性。实验实践是被设计用来分离因果机制的,故而可产生正规事件顺序,或"恒定的巧合"。但因为这些恒定的巧合是实验操作的结果,不能被等同于科学规律(否则的话,我们就得说自然规律是人类控制的产物了)。

这个分析带来了一种"分层的"世界观,把观察到的事件的"层"("经验")与"实际"的事件流(大多数都未被观察到)区别开来。被误称为"真实的"第三层(因为所有的层都在一定程度上是真实的)就是"因果机制"这一层,其趋势正是有关科学定律的陈述所要捕捉的。批判实在论可能被合理地描述成一种"深度"实在论,表征着现实在感官经验背后或下面的多层存在,有待科学实验和理论分析来发现。就这样,科学就被看成一个发现更深层的现实(基因、分子、原子、亚原子、粒子、力场、类星体等)的过程,而非对累积的经验的概括总结。对将这些科学知识应用于技术的实践的分

析，则会进一步要求将世界的本质看成是科学知识的对象。这就是它的“差异化的”特征。实验实践只是将因果机制从外来影响中分离开来（即制造“封闭系统”），而技术则涉及科学知识在“开放系统”中的应用，牵涉多种因果机制互动。在开放系统中，各种机制的相互作用会导致它们的效果被覆盖或修改，致使规则的或可预测的事件流和观察可能不会出现。

对于科学的批判实在论模型是否适用于社会科学还有不少争论。对有些社会科学哲学家而言，人类及其社会关系与自然界有很大不同，基本上无法以任何自然科学模型为基础来对其进行研究。但是，批判实在论者基本上接受一个观点，虽然他们的知识的对象和自然科学的对象有很大的区别，但社会研究仍然可以是科学的。不过，如果我们承认每种科学都必须采用适合其主旨的提问方法和解释的形式，那么用（批判实在论的）科学方法研究社会生活究竟意味着什么仍然是一个悬而未决的问题。

在 1970 年代，有些作者曾从实在论的角度提出了这个问题（Benton, 1977; Bhaskar, 1979/1998; Keat & Urry, 1975）。巴斯卡尔提出，社会科学也可以是科学——并非仅仅是姑妄言之，而是因为科学在自然和社会之间存在公认的差异，我们和这二者的关系也同样有所不同。我们之所以会认为社会结构存在，其唯一的依据是代理者的活动，取决于代理者对它们的信仰，而且只是相对持久的。关于观察者的部分同一性（the partial identity of the observer）和观察什么也都是问题：社会科学本身就是社会的组成部分，因而在社会科学中不可能建立实验式的封闭。

对巴斯卡尔（Bhaskar, 1979/1998）而言，这些差异仅仅意味着社会科学必须在其方法论中对它们予以关注——例如，开始着手研究行动者自己怎样理解他们做的事。对他来说，社会危机就好比自然科学中的实验。然而，科利尔（Collier, 1994）则认为，即使社会科学没有与自然科学中对应的实验，也不能以此为理由否定它们（社会“科学”）可以是完全是科学的：他把它们称为民智（epistemoids）。乌思怀特（Outhwaite, 1987）认为，批判实在论和批判理论相关，而本顿（Benton, 1998）则在批判实在论中看到了社会和自然科学的非还原的重新组合的基础，这一基础对理解生态问题尤为重要。

批判实在论者中有争议的另一个领域是理解与道德或政治价值之间的关系。巴斯卡尔等人认为，可以从实在论者的科学阐释中导出价值——因此（导出）“批判的”社会科学就有了可能，它面向一种人类解放的愿景。马克思对意识形态的“解释性的批判”是最著名的例子：在其他条件不变的情况下，易于让其代理产生错误观念的社会和经济关系是不合时宜的，因而应当被变革。赛耶（Sayer, 2000）反对这一点，认为只有以合乎情理的替代性解释为依据时，对现存的社会形态的批判才是正当的。

总的来说，就支持批判实在论的社会科学家而言，他们都坚定地相信，社会结构和人的能动性（human agency）之间存在差别。这两者都被视为真实、独特但也互相依赖的层面。代理者通过他们的活动复制和改变着社会结构，而社会结构既使得社会行动成为可能，也对其施加限制（Archer, 1995）。批判实在论在地理学、经济学、法学、社会学、心理学、女性主义和环境研究中还在不断发展，而其他一些领域也都受惠于批判实在论（有关批判实在论的应用的例子，请参

见 New，1996；Porter，1993）。

——Ted Benton

（杨可译　沈崇麟校）

## 参考文献


Archer, M. (1995). *Realist social theory: The morphogenetic approach.* Cambridge, UK: Cambridge University Press.

Archer, M., Bhaskar, R., Collier, A., Lawson, T., & Norrie, A. (Eds.). (1998). *Critical realism: Essential readings.* London: Routledge Kegan Paul.

Benton, T. (1977). *Philosophical foundations of the three sociologies.* London: Routledge Kegan Paul.

Benton, T. (1998). Realism and social science. In M. Archer, R. Bhaskar, A. Collier, T. Lawson, & A. Norrie (Eds.), *Critical realism: Essential readings.* London: Routledge Kegan Paul.

Bhaskar, R. (1997). *A realist theory of science.* London: Verso. (Original work published 1975)

Bhaskar, R. (1998). *The possibility of naturalism.* Hemel Hempstead, UK: Harvester Wheatsheaf. (Original work published 1979)

Collier, A. (1994). *Critical realism.* London: Verso.

Keat, R., & Urry, J. (1975). *Social theory as science.* London: Routledge Kegan Paul.

New, C. (1996). *Agency, health and social survival.* London: Taylor & Francis.

Outhwaite, W. (1987). *New philosophies of social science.* London: Macmillan.

Porter, S. (1993). Critical realist ethnography: The case of racism and professionalism in a medical setting. *Sociology, 27,* 591-609.

Sayer, A. (2000). *Realism and social science.* London: Sage.


## 批判理论（Critical Theory）

任何对社会机制的理论化和对社会关系模式的理论化如果在一定程度上与看待社会和社会关系的主流方式不同，就可以称为批判的。不过，批判理论这一词通常是用来指称与著名的社会研究的法兰克福学派及其追随者相关联的一套特殊观点。法兰克福学派的追随者对其最初的见解、愿望和议程进行了修改和扩充。

在第二任主任霍克海默（Horkheimer，1895—1973）的主持之下，法兰克福社会研究所（the Institute for Social Research at Frankfurt）将研究的重心放在了以下几个关键领域：对欧洲缺乏统一的工人阶级运动进行解释、对资本主义危机的性质和后果进行考察、对现代社会中政治和经济领域的关系进行思考、对作为政治运动出现的法西斯主义和纳粹主义进行解释、对家庭社会化进行研究、对实证主义【Positivism】和科学之间的关系不断进行批判。为了进行这一研究，需要与那些曾阻碍黑格尔、马克思以及法兰克福学派的人，例如曾质疑启蒙工程（the Enlightenment project）的叔本华和尼采进行系统的交锋，以恢复马克思主义的诺言。

综上所述，不断变化的气氛塑造了这些学者的研究。在纳粹德国，法兰克福研究所面临的威胁是其主要成员，包括阿多诺（Theodor Adorno，1903—1969）、霍克海默、马尔库塞（Herbert Marcuse，1898—1979）都移民到了美国（阿多诺和霍克海默在第二次世界大战之后回到了德国）。他们在这个研究所发现了一种自信的中产阶级的自由资本主义，它具有一种吸纳和中和无产阶级意识的显而易见的能力。这是伴随着成功和精英政治理念的个人主义意识形态的人格化。分析这些趋势及其影响需要考虑文化，直到此时，在马克思主义者的圈子中，文化

的价值一直饱受贬抑。

这些学者的研究都是跨学科的(因此预言了那些在当代更为引人注目的争论),同时伴随对理解人类关系的实证主义和诠释主义方法的批评。一方面,实证主义没有考察资本主义发展和持续的条件。另一方面,诠释主义则被认为不够精确,因为它只关注自我理解和自我意识这一过程。致使它毫无批判地接受某个给定社会内的主导的意识形态,因而无法考虑人类行动的结构因素。

于是,在启蒙工程中也加入了历史的考察。在此我们发现:与经济强权相比,个人已完全贬值,与此同时,经济强权迫使社会对自然的控制已达到了前所未有的、意想不到的高度……巨细靡遗的信息洪流和棉花糖式的娱乐(candy-floss entertainment)在指导人类的同时,也使人类变得愚蠢(Adorno & Horkheimer, 1944/1979, pp. xiv-xv)。

以此为基础,意识形态整合的过程进入了研究范畴。于是马尔库塞(Marcuse, 1968)在《单向度的人》(*One Dimensional Man*)中也提到了社会的"技术管理控制"。

这种元批评(meta-critique)注重的是下面的问题:在拒绝了诠释主义的社会理论之后,我们如何对心灵状态进行重要的分析?于是弗洛伊德作为整个精神分析和社会理论关系史中的一员,在这里登场了。这时,我们关注的已不再是理解社会世界中孤独的自我,而是没有看到自己症状成因的自我误解的人。不过,还是有问题存在。直至法兰克福派理论家时代到来之前,马克思主义者一直对精神分析嗤之以鼻,认为它会分散人们对推翻不公正经济体系这一事业的注意力。然而,精神分析毕竟还是可以被视为一种改善之举,当然,它不是解决问题之道,更别说是革命的壮举。从概念上讲,它是从个人出发的,倾向于低估社会条件和社会限制。

因此,如果弗洛伊德的工作要满足批判理论的需要,就需要进行改造。此时,利奥·洛文塔尔和弗罗姆登场了,成了批判理论这一阶段发展的中心人物。例如,弗罗姆就考察了个人和极权主义的关系。他整个一生都相信,我们爱和自由的能力都不可避免地与社会经济条件相关联。在阿多诺及其1950年代的同辈研究者有关现代人格类型结构的社会心理学研究中,对弗洛伊德的兴趣也显而易见,威廉·赖克有关性别和资本主义关系的研究中也体现出了这一点。

对批判理论来说,理论和事实之间存在着一种持续不断的互动,因而理论家都致力于探寻其命题构成与他们从中发现自身的社会背景之间的关系。有关产生什么、在什么环境下和有何种效果之间的关系的自反性【Reflexivity】,将传统理论和批判理论区别开来。此外,研究结果反馈给社会生活的议题对研究者来说不是一个"问题"。相反,批判研究的恰当性就在于它具有的鼓舞政治行动的价值。

尽管批判社会研究自认是"时代的愿望和斗争",但实际情况仍然没变,批评家们虽然承认不可能将价值和研究相分离,但还是认为试图把价值和研究混为一谈的说法站不住脚。说到底,是谁在定义何种价值才具有重要性?不仅如此,批判又是以什么为基础的?这是自由个人的自主权吗?自主权这个概念变得抽象,因为个人怎么能被假定为脱离其所属的社会关系呢?这个批判的基础太不现实,显得虚无缥缈、曲高和寡。这些批评以及其他一些批评主要源自批判理论指引下的研究实践。

在所有的当代学者中,哈贝马斯是这个传统最多产的继承者,他用大量的改造和批评处理了上述问题。这个遗产在南希·费雷泽(Fraser, 1997)和阿克塞尔·霍奈特(Honneth, 1996)以及布迪厄(Bourdieu, 2000)

那里也有所呈现，费雷泽和霍奈特对有关认知和再分配的讨论有所贡献，布迪厄则提出了“理性的现实政治”。至于福柯，他自己在《评述马克思》(*Remarks on Marx*)一文中的话可以充分地说明其观点：“法兰克福学派引发的问题还有人在不断研究。其中，权力的后果与一种西方从历史上和地理意义上定义的理性相关，而这个过程从16世纪就开始了。”(Foucault, 1991, p. 117)

——Tim May

(杨可译　沈崇麟校)

## 参考文献

Adorno, T., & Horkheimer, M. (1979). *Dialectic of enlightenment* (J. Cumming, Trans.). London: Verso. (Original work published 1944.)

Bourdieu, P. (2000). *Pascalian meditations* (R. Nice, Trans.). Cambridge, UK: Polity.

Comstock, D. E. (1994). A method for critical research. In M. Martin & L. C. McIntyre (Eds.), *Readings in the philosophy of social science* (pp. 625-639). Cambridge: MIT Press.

Foucault, M. (1991). *Remarks on Marx: Conversations with Duccio Trombadori* (R. J. Goldstein & J. Cascaito, Trans.). New York: Semiotext(e).

Fraser, N. (1997). *Justice interruptus: Critical reflections on the "postsocialist" condition.* London: Routledge Kegan Paul.

Honneth, A. (1996). *The struggle for recognition: The moral grammar of social conflicts* (J. Anderson, Trans.).Cambridge: MIT Press.

Marcuse, H. (1968). *One dimensional man: The ideology of industrial society.* London: Sphere.

Morrow, R. A. (with Brown, D. D.). (1994). *Critical theory and methodology.* London: Sage.

Wiggershaus, R. (1995). *The Frankfurt School: Its history, theories and political significance* (M. Robertson, Trans.).Cambridge, UK: Polity.

# 临界值(Critical Values)

在假设检验【Hypothesis Test】中，我们确定数据是否极端到足够对虚无假设【Null Hypothesis】产生怀疑。如果数据是极端的，我们就拒绝虚无假设；如果数据是适中的，我们就接受虚无假设。在极端和适中的数据之间、接受和拒绝虚无假设之间的阈值称为临界值。

假设我们掷一枚硬币10次。一个显而易见的虚无假设是这枚硬币是公正的，或出现面朝上和面朝下的可能性是相等的。如果这个虚无假设是真的，那么我们可能将掷出接近5次的面朝上。如果面朝上的数目接近5(适中)，一个明智的检验将会接受虚无假设；而如果面朝上的数目与5相去甚远(极端)，检验就会拒绝虚无假设。例如，如果我们掷出了9次或更多的面朝上，或者我们掷出了1次或更少的面朝上，那么一个检验就可以拒绝虚无假设。对这个检验而言，1和9就是临界值。

临界值与显著水平【Significance Level】是相关的。上面的检验，其显著水平是0.021 5。在上面这一虚无假设下，面朝上的次数至少会像临界值那样极端的概率为0.021 5的(这个概率是用二项分布【Binomial Distribution】计算的)。我们可以用这个临界值来计算显著水平，这就像我们在这里计算的一样。然而更为普遍的做法是，先选择显著水平，然后再计算临界值。

用临界值进行检验，等价于用$p$值进行检验。这就是说，因为极端的数据而拒绝一个假设，等同于因为一个小的$p$值而拒绝一个假设。例如，掷10次硬币产生了10次面

朝上,10 次是比临界值 9 次更为极端的数目,这样我们将用临界值来拒绝虚无假设。但是如果我们用双尾【Two-Tailed】的 $p$ 值,也可以拒绝这个虚无假设。10 次面朝上的 $p$ 值是 0.002,它小于显著水平 0.021 5。诚如本例所示,基于 $p$ 值的决定,总是与基于临界值的决定是一致的。

——Paul T. von Hippel

(沈崇麟译校)

## Cronbach $\alpha$(Cronbach's Alpha)

Cronbach $\alpha$ 是一种用于评估含有一个以上成分(如题项、评级)的综合量度的应答的一致性的内在信度【Internal Reliability】估计量。$\alpha$ 系数是使用最为广泛的信度估计量,它是由克龙巴赫于 1951 年首先引进的。根据《社会科学引文索引》(*Social Sciences Citation Index*),在 1951—2002 年,这篇文章被其他作者引用了 4 912 次。那些在文章中提到了 Cronbach $\alpha$,但没有引用原文的、出版的和未出版的文章数很可能是一个天文数字。

Cronbach $\alpha$ 和标准化的 Cronbach $\alpha$ 可用公式表示为

$$\alpha = \left(\frac{k}{k-1}\right)\left(1 - \frac{\sum_{i=1}^{k} s_i^2}{s_t^2}\right)$$

而标准化的公式则是

$$\alpha = \frac{k\bar{r}}{1 + \bar{r}(k-1)}$$

式中的 $k$ 是综合量度中的成分数,$s_i^2$ 是成分 $i$ 的方差,$s_t^2$ 是量度上的总分数的方差,而 $\bar{r}$ 则是平均的成分间的相关。在需要计算成分的标准化分数,而不是原始的成分分数时,就要使用标准化的 Cronbach $\alpha$。Cronbach $\alpha$ 的精度取决于成分间相关的标准误差,其计算公式为

$$\frac{s_r}{\sqrt{[0.5 \times k \times (k-1)] - 1}}$$

式中的 $s_r$ 是成分间相关的标准差。在所有其他方面相等时,如果在成分之间相关之中的变差变大,那么 Cronbach $\alpha$ 的精度便会降低。然而,应该注意的是,Cronbach $\alpha$ 的大小并非标准误差的大小的函数。可能会有两套 $k$ 个成分,在成分之间相关之中的变差量虽然不同,但 Cronbach $\alpha$ 系数却相似,因为它们取决于平均的成分之间相关。

从概念上讲,是 Cronbach $\alpha$,而非标准化的 Cronbach $\alpha$ 才是由一个基于所有成分的公共因子解释的综合量度的方差比例的下限估计值。尽管这个陈述说明 Cronbach $\alpha$ 不应该是负的,且只能在 0~1 变化,但实际上观察到一个负的 Cronbach $\alpha$ 并不是不可能的。表 1 列出了一个七个题项的量度施于 10 个人一组的测量结果。每个题项有三个从 1~3 的答案类别。阿尔法系数是负的,而它的绝对值甚至大于 1。

Cronbach $\alpha$ 可能会受到回答人和成分的特性的影响。在每件事都相等时,如果回答人是同质的(即在一个量度评估的特性上是相同的),那么 $s_t^2$ 会趋于下降。正如从上面的公式看到的一样,一个小的 $s_t^2$ 将产生一个小的阿尔法系数。其次,随着高质量成分数的增加,Cronbach $\alpha$ 也趋于变大。然而这些正向特征,在成分数太大,致使 $\frac{k}{k-1}$ 接近 1 时,就不再那么突出。注意,如果一个量度的成分是异质的,这就意味着潜在于成

分之中的因子或维度在一个以上。Cronbach $\alpha$ 将会低估这个真的信度。此外，皮得森（Peterson，1994）的元分析【Meta-Analytic】发现指出，在一个量度由少于四个题项、每个题项只有两个答案类别的题项组成时，Cronbach $\alpha$ 将会低估真信度。必须强调的是，高的 Cronbach $\alpha$（如 0.8）并非必定意味着潜在的成分因子或维度必定只有一个。高内在一致性是单维性（unidimensionality）的必要条件而非充分条件。还有一个应该注意，但却常常为人们所忽略的问题是，Cronbach $\alpha$ 并非对所有复合量度的信度估计都适合。例如，在一个复合量度的成分是一种构造的因果指标时，就不宜用 Cronbach $\alpha$ 来做信度估计。举一个例子，假如有一张描述社会活动的构建的清单，上面列出的行为有看电视、看书、滑雪、访友和下棋。而这些活动不能被看成是一些平行的成分。所以 Cronbach $\alpha$ 是不能为这一量度提供精确的信度估计的。

**表 1 负的 Cronbach $\alpha$ 例证**

| 人 | 题项 1 | 题项 2 | 题项 3 | 题项 4 | 题项 5 | 题项 6 | 题项 7 | 总计 |
|---|---|---|---|---|---|---|---|---|
| A | 1.00 | 2.00 | 2.00 | 2.00 | 1.00 | 2.00 | 2.00 | 12 |
| B | 2.00 | 1.00 | 2.00 | 3.00 | 2.00 | 3.00 | 2.00 | 15 |
| C | 1.00 | 2.00 | 1.00 | 2.00 | 2.00 | 2.00 | 2.00 | 12 |
| D | 2.00 | 3.00 | 2.00 | 1.00 | 2.00 | 1.00 | 2.00 | 13 |
| F | 1.00 | 2.00 | 3.00 | 2.00 | 1.00 | 2.00 | 3.00 | 14 |
| G | 3.00 | 1.00 | 2.00 | 3.00 | 2.00 | 1.00 | 2.00 | 14 |
| H | 1.00 | 2.00 | 2.00 | 2.00 | 2.00 | 2.00 | 1.00 | 12 |
| I | 2.00 | 1.00 | 2.00 | 1.00 | 2.00 | 1.00 | 2.00 | 11 |
| J | 1.00 | 2.00 | 1.00 | 2.00 | 3.00 | 2.00 | 1.00 | 12 |
| K | 1.00 | 3.00 | 2.00 | 1.00 | 2.00 | 2.00 | 2.00 | 13 |
| 方差 | 0.45 | 0.49 | 0.29 | 0.49 | 0.29 | 0.36 | 0.29 | 1.36 |

$$\alpha = \left(\frac{k}{k-1}\right)\left(1-\frac{\sum_{i=1}^{k} s_i^2}{s_t^2}\right) = \left(\frac{7}{6-1}\right) \times \left(1-\frac{2.66}{1.36}\right) = -1.12$$

$$\text{标准化}\alpha = \frac{k\bar{r}}{1+\bar{r}(k-1)} = \frac{7\times(-0.080\,2)}{1+(-0.080\,2)\times(7-1)} = -1.08$$

——Peter Y. Chen
Autumn D. Krauss
（沈崇麟译校）

## 参考文献

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, *78*, 98-104.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*, 297-334.

Osburn, H. G. (2000). Coefficient alpha and related internal consistency reliability coefficients. *Psychological Methods*, *5*, 343-355.

Peterson, R. A. (1994). A meta-analysis of Cronbach's coefficient alpha. *Journal of Consumer Research*, *21*, 381-391.

## 跨文化研究(Cross-Cultural Research)

从人类学的意义上来说,跨文化研究可以被定义为对不同文化的任何描述或比较。跨文化研究也可在系统比较的意义上使用,其明确的目标在于回答那些有关文化差异的发生率、分布和成因的问题(Ember & Ember, 2001)。尽管它对于确认各种文化中重复出现的主题和模式很有用,但在对文化的相同点做概括时可能因为心理定式的危险而有一定的风险。模式可以给跨文化的理解提供整体的结构,但这些模式也可能会遮蔽每种文化内的极大的个体差异。

无论其目标是描述一种区别于研究者文化的独特文化,还是进行跨文化的比较,在进行跨文化研究时总是会出现许多问题。种族中心主义也都用自己的文化作为评判别人的透镜或标准,对于理解另一种文化经验的一种障碍。文化相对主义重视文化是差异性,强调要从其他文化自身的环境中来理解其他文化中的人。

在跨文化比较中,样本内的差异性,以及非文化维度上的差异性可能会对比较研究的效度【Validity】产生危害(Hujala, 1998)。在一项有关美国和芬兰的教师的想法的研究中,研究者希望用国家的政治边界来将文化定义为他们的自变量。但是,芬兰的教师和美国的教师在其他的重要维度上存在着差异。芬兰的教师比美国的教师所受的培训更多,而这种差异可能解释一定社会的教师想法的差异,而非文化想法的差异。戈保罗-麦克尼科尔和阿穆尔-托马斯(Gopaul-McNicol & Armour-Thomas, 2002)提出,研究者应考察文化和阶级、文化和民族、文化和语言之间的关系,以更好地理解如何观察文化群体之内和之间的有意义的差别之处。他们提出了与一个国家的优势文化和次优势文化相关的问题,以及与每个群体有联系的认识论【Epistemological】假定。经由主流文化群体发展的研究方法和数据收集工具,可能会造成对非优势群体的个人经验相对忽视。

跨文化研究的团体对跨文化的测量可以发展到何种程度有一番争论,反映出在此问题上的不同立场(Evans, 2000)。其中一个立场是,测试可用适当的语言进行翻译,由本国测试人员实施,提供熟悉的内容,从一种文化转换成另一种文化。文化心理学家则持有一种不同的立场,因为他们强调符号文化的重要性,也就是说,强调共享的价值和意义、认知和交流的重要性(Greenfield, 1997)。因此,只有在具体问题的具体答案的取值或价值达成一致的意见、在不同文化中相同题项的含义必须相同,且对这些研究工具有上佳的语言翻译时,一个测试才能被用于另一个文化之中。此外,对相关信息的定义也必须普遍一致,而且和陌生人交流也必须为人所接受。

为了得到跨文化研究的有效数据，人们必须确定概念的地方定义(Evans，2000)。在西方国家与在肯尼亚乡下小村对智力的测量可以作为一个例子来说明概念的地方定义是多么重要。在肯尼亚乡下，村民描述孩子智力的术语是他们能说出多少动物和树木的特点、用途和名称，或者在指定的距离内，孩子们能把凳子或者食物这类东西搬多远。这些就是村民们在孩子要入学时用来判断他们成长是否正常的标准。

在研究项目中文化表达对进行可靠的文化研究至关重要。这个原则适用于多文化的国家内以及跨国的研究。这需要你从研究过程一开始就全身心地投入。正在进行的教育成就评估国际协会的学前教育研究(Preprimary Study of the International Association for the Evaluation of Education Achievement，Olmsted & Weikart，1994)就是遵循这个原则的一个跨文化研究的例子。这项研究被设计用来调查各种不同幼儿期经验的性质及其对幼儿的长期发展有什么影响。这个研究的过程和工具是由最初参与这个研究的各国(即比利时、中国、芬兰、德国、中国香港、意大利、尼日利亚、葡萄牙、西班牙、泰国、美国)学者所组成的跨文化的研究团队开发的。整个团队来决定研究要包括哪些内容，以及能最好地表现他们的文化的例子和措辞。

在跨文化的研究中，测量的效度可能会受到研究工具的翻译以及与陌生人或长者相处的文化规范的影响(Evans，2000)。即使是精确的语言翻译也仍然有可能会带来误解，因为词汇往往会附带多种意义。翻译应该不仅是字面的翻译，还应该是含义的翻译。和孩子一起做研究有可能会受到文化规范的影响，尤其是在成年人普遍不询问孩子的意见，或者不主张孩子在成人在场时说话的情况下。在这种情况下，研究者可考虑做一个表现性评估，这种评估并不是直接提问。

——Donna M. Mertens

(杨可译　沈崇麟校)

## 参考文献

Ember，C. R.，& Ember，M. (2001). *Cross-cultural research methods.* Walnut Creek，CA：AltaMira.

Evans，J. L. (2000). *Early childhood counts.* Washington，DC：World Bank.

Gopaul-McNicol，S.-A.，&Armour-Thomas，E. (2002). *Assessment and culture：Psychological tests with minority populations.* San Diego：Academic Press.

Greenfield，P. (1997). You can't take it with you：Why ability assessments don't cross cultures. *American Psychologist*，*52*(10)，1115-1124.

Hujala，E. (1998). Problems and challenges in cross-cultural research. *Acta Universitatis Ouluensis*，*35*，19-31.

Olmsted，P.，& Weikart，D. P. (Eds.). (1994). *Families speak：Early childhood care and education in 11 countries.* Ypsilanti，MI：High/Scope Press.

## 交叉滞后模型(Cross-Lagged)

交叉滞后模型广泛用于追踪数据【Panel Data】，或同一个体或单位在不同时间多次收集的数据，为 $X$ 和 $Y$ 之间的因果关系的方向提供证据，并估计每个变量在另一个变量上的因果【Causal】效应的强度。交叉滞后模型涉及在某一期追踪调查测量的每个变量和下一期调查的变量之间的相关【Correlation】或回归系数【Regression

Coefficients】的估计和比较问题。这样的模型潜在的逻辑是，如果 $X$ 是 $Y$ 的一个原因，那么在时间 1 量度的 $X$ 应当和时间 2 量度的 $Y$ 有关，这就是说，$X_1$ 和 $Y_2$ 之间的交叉滞后关系应当是非零的。不仅如此，如果 $X$ 对 $Y$ 的影响大于 $Y$ 对 $X$ 的影响，那么，$X_1$ 和 $Y_2$ 之间的交叉滞后关系，应当强于 $Y_1$ 和 $X_2$ 之间的关系。

如果变量以平均的离差式表示，那么截距项便可省略。

一个两期的追踪调查的交叉滞后模型如图 1 所示。第二期调查的变量 $X_2$ 和 $Y_2$ 被假设取决于它们的第一期调查的值、其他变量的滞后【Lagged】值和误差项 $U$；第一期的 $X$ 和 $Y$ 之间的相关系数由 $\rho X_1 Y_1$ 表示，而方程的误差【Error】项之间的相关则由 $\rho U_1 U_2$ 表示。如果变量以平均离差的形式表示，那么省略了截距项的两个方程可写成

$$Y_2 = \beta_1 X_1 + \beta_2 Y_1 + U_1 \qquad (1a)$$

$$X_2 = \beta_3 Y_1 + \beta_4 X_1 + U_2 \qquad (1b)$$

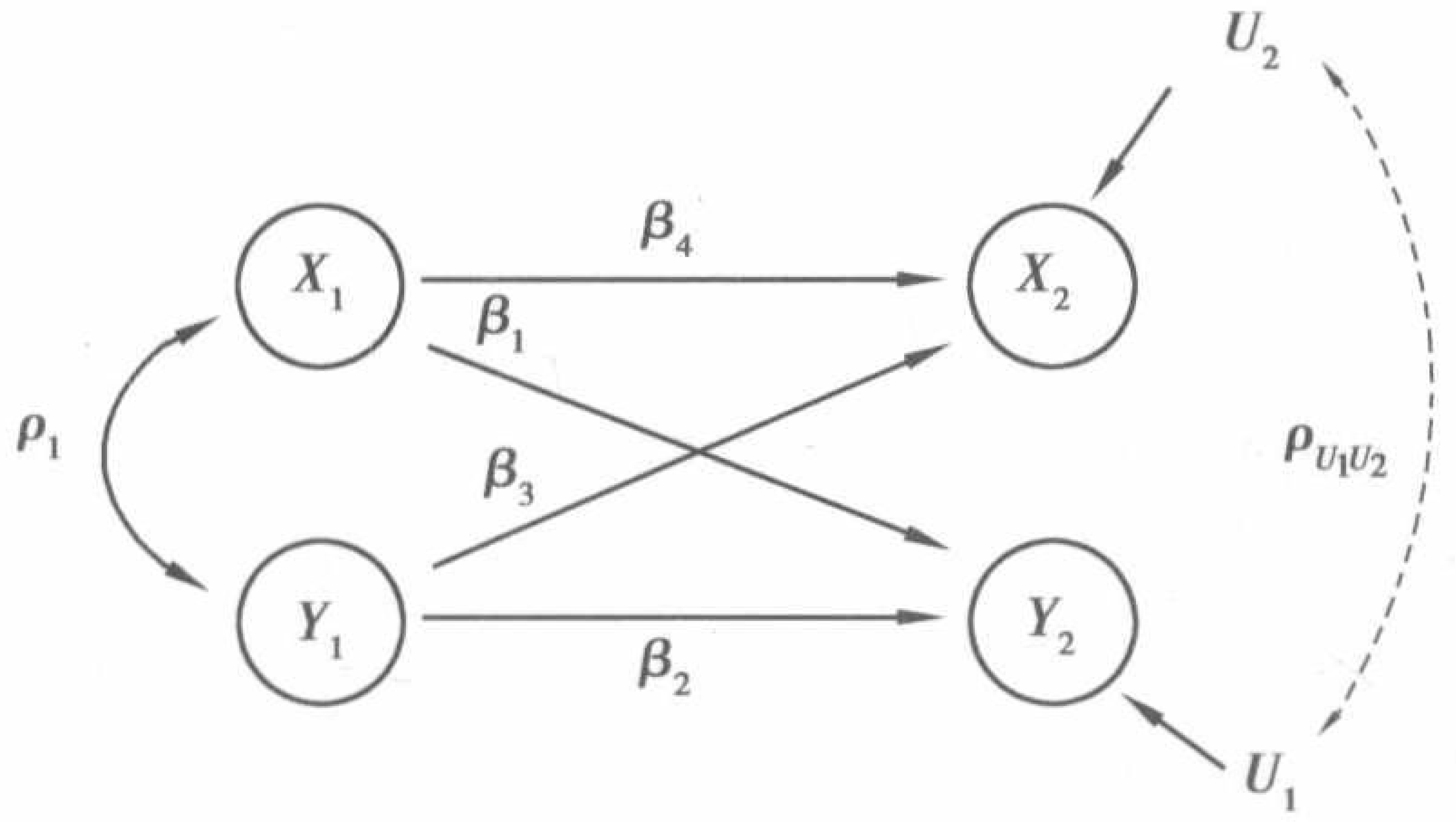



**图 1　交叉滞后效应的两期追踪模型**

我们在最初就已论证，两个变量之间的因果方向可通过比较两个交叉滞后的相关来确定，也就是看究竟 $\rho X_1 Y_2$ 大于还是小于 $\rho Y_1 X_2$ 来确定。然而，很多学者指出，交叉滞后相关是图 1 中因果效应 $\beta_1$（或 $\beta_3$），以及每个变量的历时稳定性（如 $\rho X_1 X_2$ 或 $\rho Y_1 Y_2$）及 $X$ 和 $Y$($\rho X_1 Y_1$)之间的初始相关的函数。因此，尽管现在仍有人在检验 $X$ 和 $Y$ 之间的假关联的某种模型中使用交叉滞后相关分析，但在因果方向和强度问题上，使用更为广泛的则是回归【Regression】和结构方程模型【Structural Equation】(Kenny, 1979)。

在交叉滞后追踪模型的因果推论中存在几个重大的风险。首先，$Y_1$ 和 $X_1$ 的量度误差【Measurement Error】将会降低 $X$ 和 $Y$ 的历时稳定性的估计值的精度，因此，也会导致交叉滞后估计值的偏倚(Kessler & Greenberg, 1981)。其次，正像在纵向【Longitudinal】数据常会遇到的一种情况一样，误差项 $U_1$ 或 $U_2$ 可能会自相关【Autocorrelated】，从而导致 $Y_1$ 和 $U_1$ 及(或) $X_1$ 和 $U_2$ 之间的非零关系。更为重要的是，交叉滞后模型可能会产生错误的结果，因为 $X$ 和 $Y$ 之间真的因果滞后大大短于测量之间的时间期(Plewis, 1985)。在遇有这样的情况时，恰当的做法是设定一个“同步(synchronous)”或“同时(cotemporal)”效应模型，在这种模型中，假设 $X_2$ 和 $Y_2$ 在同一时期相互影响。兼有交叉滞后和同步效应的模型也是可以设定的。这种模型可用工具变量【Instrumental Variable】、相关的程序或在多期调查中通过在跨期的参数上设置等式或其他的约束来设定和估计。

——Steven E. Finkel

（沈崇麟译校）

## 参考文献


Kenny, D. A. (1979). *Correlation and causality*. New York: John Wiley.

Kessler, R., & Greenberg, D. (1981). *Linear panel analysis*. New York: Academic Press.

Plewis, I. (1985). *Analysing change: Measurement and exploration using longitudinal data*. Chichester, UK: Wiley.

## 横截面数据(Cross-Sectional Data)

横截面数据来自一种每个参与者的所有观察都在几乎同一个时点收集,且时间并非研究设计【Research Design】中的要素的调查研究。所有观察都被认为是同时作出的,尽管实际上这绝不会是真的。即使在单个访谈中,个人也是按时间顺序来完成一个一个题项的,而在一个研究中,参与者一般是不会在同一时间提供数据的。然而,这样一种设计的时间期限一般都比较短,因而我们假定研究期间的观察时间有只会产生误差的随机【Random】效应。这与纵向【Longitudinal】设计不同,在纵向设计中,观察是在不同的时点进行的,以了解变量是否会因为在研究过程中发生的那些事件而有所不同。

横截面一词常被用来描述问卷【Questionnaire】研究。最简单的横截面设计是单源横截面设计(the single source cross-sectional design),在这种设计中,参与者提供有关自己所有的数据,且所有的题项和量表都在单独一次(访谈)单元内完成。多源设计增加了一些未包含在参与者问卷中的数据【Data】,因而有可能包括给其他人(如同事或家庭成员)的问卷,观察员提供的数据或档案数据。另一种方法是将一份长问卷分成在两个不同访谈单元实施的两个部分,但并没有要做一些有关时间问题的推论的想法。

我们有可能不用问卷调查和在抽样单位并非一个人的研究中获取横截面数据。例如,我们可以看看美国人口普查中的不同变量之间的关系,而美国人口普查的所有数据则被认为是在同时收集的。

横截面数据对于确定两个变量【Variable】是否有关系这样的问题效率很高。它们最大的局限是不能作出因果【Causal】结论(Shadish, Cook, & Campbell, 2002; Spector, 1994)。不将时间纳入设计,我们就无法确定两个关联的变量之间的因果方向,或是否有另外的可解释观察到的关系的变量。为了达到这一目的,我们必须采用纵向设计或做一个实验【Experiment】。

——Paul E. Spector

(沈崇麟译校)

### 参考文献

Shadish, W.R., Cook, T.D., & Campbell, D.T.(2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton-Mifflin.

Spector, P.E.(1994). Using self-report questionnaires in OB research: A comment on the use of a controversial method. *Journal of Organizational Behavior*, *15*, 385-392.

## 横截面设计(Cross-Sectional Design)

横截面设计与一种假定数据【Data】在同一时点收集的数据收集法有关。其目的是得到一个群体的一张"快照"或照片。我们假设数据能代表个人、群体、单位、行为或其他的分析单位,而数据则可用一系列程序来收集。实际上,横截面设计一词常与调查【Survey】一词互换,而在大多数时候都是指那些通过对代表自己、家庭、机构或其他的

社会实体的个人进行的访谈或问卷调查。

尽管不是硬性规定的，但是横截面设计确实常常假定研究的群体或总体的特性是异质的，这就是说，在研究总体【Population】中的许多年龄、行为和观点不同的人是被代表的。（横断面设计）的目的是设法恰当地“代表”群体内的多样性。由异质性假定可推出以下假定：在分析发现和下结论时，适当地代表和推论较大的总体或群体是必要的，因为我们研究的对象都是从那里抽取的。正因如此，我们必须对研究对象的抽取方法予以特别的关注。如果在横截面研究涉及的范围内，数据并非从整个感兴趣的总体抽取的，那么了解研究对象是如何抽取的这一问题是十分重要的。概率【Probability】或系统抽样【Systematic Sampling】策略通常是横截面设计的首选。因为横截面设计的数据是在同一个时点收集的，因此分析的重点是考察单个变量【Variable】的频数分布【Frequency Distribution】和两个或多个变量之间的关联【Association】。研究者有时把描述性和解释性的【Explanatory】横截面研究加以区别，因为超越描述的解释性研究旨在考察某些事物是如何影响或导致变化的（如探讨遭受自然灾害是否会增加心理压力），并试着就这些问题得出一些结论。

横截面设计可以与群体比较或个案控制【Case Control】设计、纵向【Longitudinal】、前瞻性或追踪【Panel】设计和实验【Experimental】或准实验【Quasi-Experimental】设计相比较。纵向、实验和群体比较设计的目的在于加大预测、调查因果【Causal】链和“解释”某些事物为什么和怎么样发生的能力，而横截面设计却能更好地描述两个变量之间的关系。这些设计（如实验设计）不太关心外在效度【External Validity】，或是否需要推广到一个更大的总体这样的问题，因而它更关心内在效度【Internal Validity】，或需要精确地测量变量或建构之间的关系。

社会学家早期做的大多数研究都是有关怎样设计横截面研究的。保罗·拉扎斯菲尔德是对早期的横截面研究作出很大贡献的主要研究者之一。而莱斯莉·基什则对横截面研究的抽样问题作出了很大贡献（Kish，1965；Lazarsfeld，1958）。在计算机技术取得长足的进步之前，拉扎斯菲尔德对横截面数据的因果推论法作出了特别的贡献。他在精细模型【Elaboration Model】方面的工作，后来被莫里斯·罗森堡（Rosenberg，1968）进一步拓展。

——Linda B. Bourque

（沈崇麟译校）

## 参考文献


Aday, L. A. (1996). *Designing and conducting health surveys: A comprehensive guide* (2nd ed.). San Francisco: Jossey-Bass.

Kendall, P. L., & Lazarsfeld, P. F. (1950). Problems of survey analysis. In R. K. Merton & P. F. Lazarsfeld (Eds.), *Continuities in social research: Studies in the scope and method of "the American soldier"* (pp. 160-176). Glencoe, IL: Free Press.

Kish, L. (1965). *Survey sampling*. New York: John Wiley.

Lazarsfeld, P. F. (1958). Evidence and inference in social research. *Daedalus*, *87*, 120-121.

Rosenberg, M. (1968). *The logic of survey analysis*. New York: Basic Books.

## 交叉表(Cross-Tabulation)

交叉表是一种显示两个变量之间的关系【Relationship】的表。它也称为交互表或列联表【Contingency Table】。最简单的交叉表是一张 2×2 的表,它包含四个格【Cells】[每一个对应(变量的)值的一个可能组合]。假设一个变量是性别(男=1,女=0),而另一个变量是是否投票(是=1,否=0)。假定性别是自变量【Independent Variable】,它可被置于表格的顶部(作为列变量)。是否投票假设是因变量【Dependent Variable】,可将它置于表格的侧面(作为行变量)。原始频数和百分数频数可置于每一格中;例如,格 A 中的数字可以是 214(55%)个投过票的男性。变量的类别越多,表中的格子数也就越多。此外,分析者也可在控制第三个变量的同时,观看一张交互表中两个变量之间的关系。例如,我们可以分别在白人和非白人的子样本内,查看性别与是否投票的表。

——Michael S. Lewis-Beck

(沈崇麟译校)

*也可参见列联表【Contingency Table】。

## 累计频次图(Cumulative Frequency Polygon)

累计频次图或累计曲线是频次图的一个变种。虽然二者都用于描述一组较大的定量数据,但区别在于累计频次图显示的是 $y$ 轴上的累计频次,它显示的频次既可以是绝对数(计数),也可以是相对数(比例)。累计频次对于了解值的数目或比例位于某个给定的值之上或之下是很有用的。

在图 1 的样本累计频次中,我们很容易就判断身高矮于 68 英寸(1 英寸=2.54 厘米)、等于 68 英寸和高于 68 英寸的男性的数目。将 $y$ 轴上的值变为比例(200/1 000=0.20、400/1 000=0.40 等),就会产生矮于、等于和高于 68 英寸的男子的比例。因为在频次图中,$y$ 轴产生的是单独一类内的观察频次(绝对或相对)量度,而非跨类别累计的量度,所以要在一张频次图中得到同样的信息并不是那么容易的。

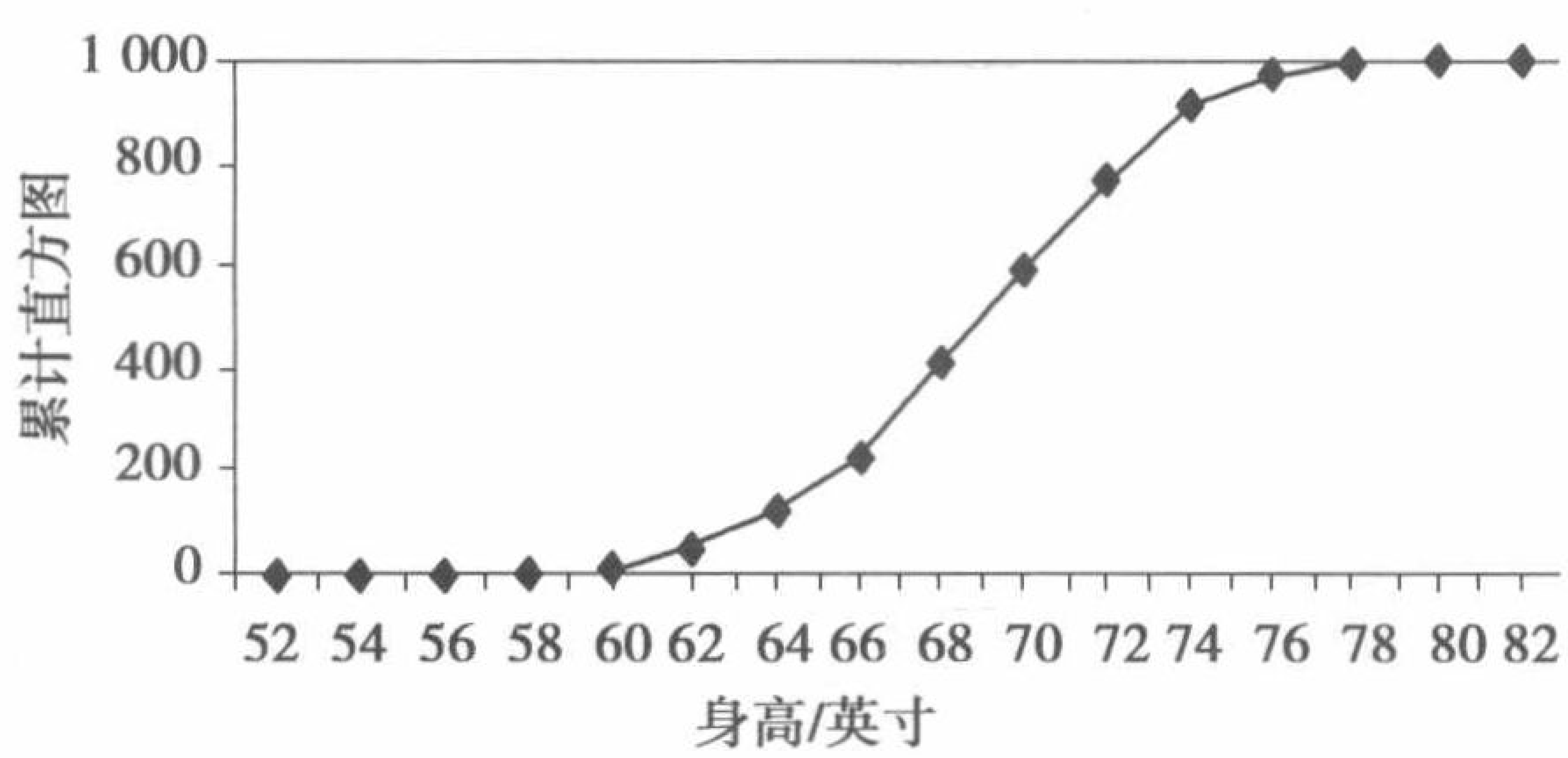



**图 1 身高在 52~82 英寸的成年男性,以 2 英寸为间隔的身高累计频次图**

为了创建一张累计频次图,研究者应先将数据由高到低排序,然后将它们归入连续的间隔。每个间隔的上限标示在一张图的 $x$ 轴上,$y$ 轴则提供累计频次的量度。它显示数据值落在或低于每一上限的数目或比例。对于绝对频数而言,尺度从最小值 0 到最大

值，它等于数据组量度的数目。对于相对频次而言，尺度则从 0 到 1.00。在数字范围内，一条条直线连接标志累积频次的点。

在数据的间隔数与数据集中不同值的数目不相等时，常用累积频次图来表示定量数据。对于连续数据而言，情况总是这样的。在数据离散时，有可能在 $x$ 轴上标示出数据的每一个值。这时应使用累积直方图，而不是累积频次图。二者的主要区别在于累积频次图的点是用直线连接的，允许在非上限处（如在图 1 的 67 英寸处）插入 $x$ 轴值的频次。当数据除了那些在 $x$ 轴上标示的值之外没有（其他）值的时候，累积频次只需简单地用直条（或列）标示即可。

有些统计学家用累计频次图一词特指那些显示值的绝对频次的图，而用累计相对频次图一词来指那些显示比例的图。累计频次图的一种备择用法，即本词条采用的用法，则同时指这两种类型的图，即计数的和比例的累计频次图。

——Kenneth O. McGraw

（沈崇麟译校）

## 曲度（Curvilinearity）

社会科学中的变量可以展示彼此间的一种曲线关系，这意味着它们间的变化取决于每个变量所取得的数值。通常变化先以自变量【Independent Variable】的值表达，然后才是变化对因变量【Dependent Variable】的影响。如果将斜率【Slope】的变化标示在一条线上，它将呈一种曲线状，就像一张有着沿 $x$ 轴的自变量和沿 $y$ 轴的因变量的二维图所显示的一样。处理曲线关系对普通最小二乘【Ordinary Least Squares，OLS】分析是至关重要的。

在这里，将列举一个来自政治科学的有关这种现象的例子，该例是关于经济发展［人均国民生产总值（GNP）］与跨国家的民主之间的关系。随着那些十分贫穷的国家的财富的增加，它们可能也会引人注目地提升自己的民主表现水平（democratic performance）。那些财富水平很低的国家可望在政治绩效方面有很大的直线上升（不言而喻，这种关系并不完美，因为没有一种单一的因果解释能完全令人满意地对社会科学现象作出解释）。然而民主的初始的剧烈增长，会随着国家在经济上越来越发达而趋于温和。图 1 所示的这种曲线效应是显而易见的。关系的斜率从财富低水平时的陡峭，转向财富高水平时的几乎为零，大多数国家都聚集在高民主值的周围。

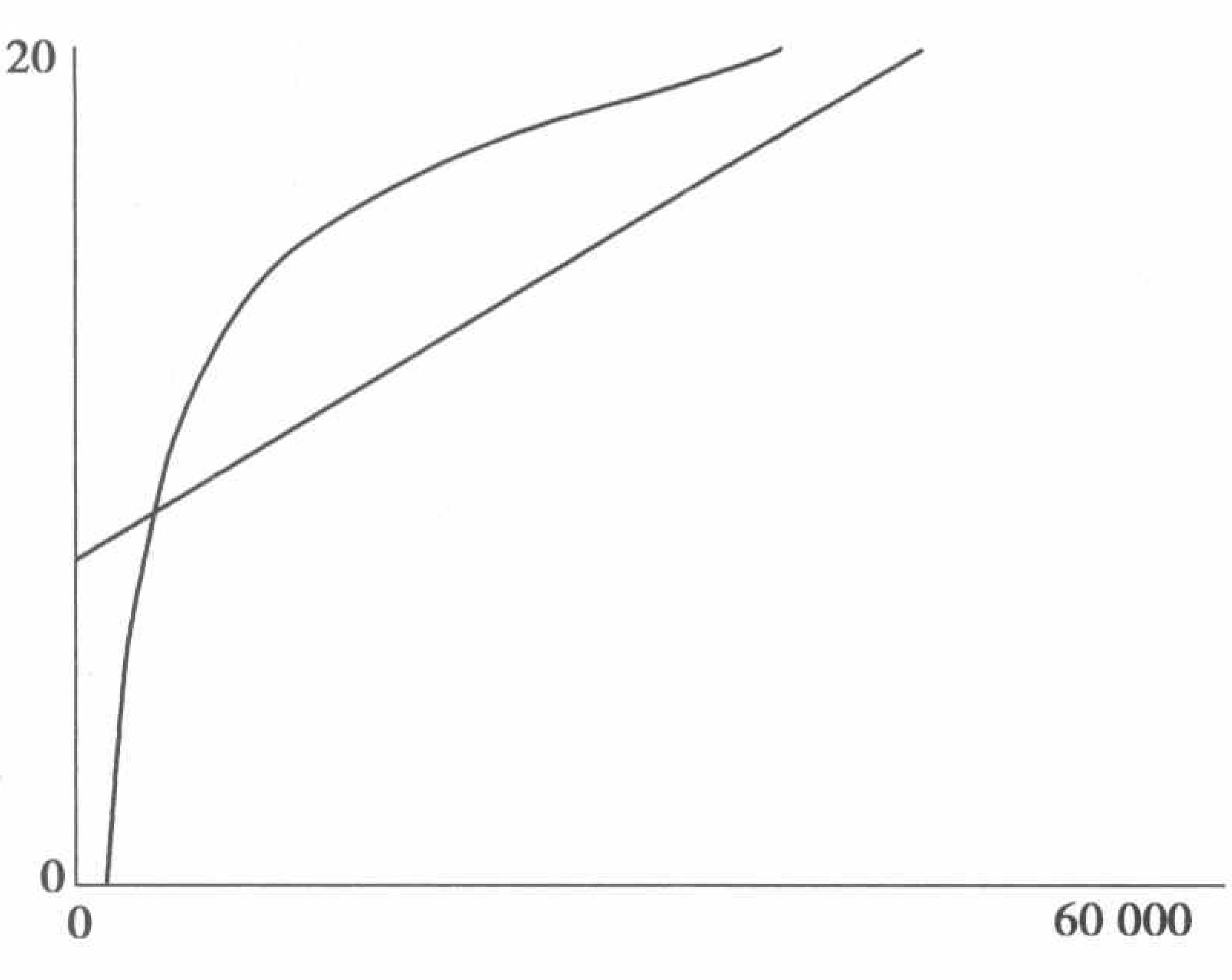



**图 1　经济发展与民主的曲线（与线性）关系**

对旨在产生估计值的 OLS 回归而言，若要产生最优的线性无偏估计值（BLUE），就必须使回归的假定得到满足。而在这些假定中，有一个便是自变量必须和因变量有线性关系。在人均 GNP 和民主的关系这一

例子中,这个假定显然被违反了。由此得到的 OLS 估计值将是有偏的估计值。

为了解决 OLS 回归中的曲度问题,分析者可以设法进行补救,所有补救方法的要求都在于通过变量的算术变换来满足线性假定。社会科学的分析者常使用三种转换法中的一种。最常用的是对数转换,这种转换既可用自然对数【Logarithm】,也可用以 10 为底的对数(无论用哪一种,作用均相同)。对数转换试图将紧紧堆积的数据变得离散,从而"线性化"散点。对数转换可线性化经济发展与民主的曲线关系。

虽然图 1 并没有显示点的实际散布,但是它显示曲线捕捉的点的分布形状优于直(线性)线。统计上显示,OLS 回归模型的拟合的改进是显著的。线性设定的 $R^2$ 统计量的值是 0.27,而对数设定(取 $X$ 的对数)的 $R^2$ 的值则是 0.30。[数据集有 $N=130$ 个国家,排除了两类异常值【Outlier】国家,OPEC(石油输出国组织)成员国和那些石油出口的收入占国家出口收入 50%以上的国家。]

另外两种常用的转换是抛物线转换和双曲线转换。如果关系的图形呈倒转的 U 形曲线,那么我们建议采用抛物线转换。这种转换需要取自变量的平方,并将它与未变换的变量一起纳入 OLS 回归模型。如果因变量有一个阈值,那么我们建议使用双曲线转换。为了实现这种转换,研究者应该取自变量 $X$ 的倒数 $1/X$。在未变换的变量有曲线关系时,这三种变换将有助于满足线性假定。

——Ross E. Burkhart

(沈崇麟译校)

## 参考文献


Burkhart, R. E., and Lewis-Beck, M. S. (1994). Comparative democracy: The economic development thesis. *American Political Science Review. 88*, 903-910.

Draper, N. R. (1998). *Applied regression analysis* (3rd ed.). New York: John Wiley.

Jackman, R. W. (1973). On the relationship of economic development to political performance. *American Journal of Political Science*, *17*, 611-621.

Kruskal, J. B. (1968). Transformation of data. In D. L. Sills (Ed.), *International encyclopedia of the social sciences* (Vol. 15, pp. 182-192). New York: Macmillan.

Studenmund, A. H. (1992). *Using econometrics: A practical guide* (2nd ed.). New York: Harper-Collins.

# D

## 研究中的危险(Danger in Research)

研究者在进行实地调查时,可能会遇到各种各样的风险。这些风险可能包括意外事故、遭监禁或遇袭、生病、遭到骚扰和谩骂等。最有可能遇到这些风险的是那些工作在边缘地区、暴力社会冲突场合,以及和某些边缘群体或从事高危职业的人群打交道的研究者。女性的研究者可能会遇到性侵犯或性骚扰。实地调查产生的危险会阻止研究者进行特定主题的调查,或在特定地区进行的工作。例如,涉及的危险可能会妨碍对发生在一些国家的暴力冲突进行研究。调查现场的风险也可能与现场各种情况的伦理道德和政治有着很强的联系,并会对专业实践产生影响。

概括地讲,危险有两类,即环境的(ambient)和情境的(situational)。环境危险是一种本来可以避免,但因进行调查而不能避免的危险。例如,在研究有关暴力社会冲突的执法研究中,研究者可能必须随同警官一起到那些极易受到攻击的地区去。情境危险是一种因研究者的存在或活动而引起的那些情境中的人的攻击、敌意或暴力侵犯的危险。例如,人们已经知道,吸毒者在受某种毒品的影响时,可能会对研究者发起攻击。在危险境况中工作的研究者,最易受到攻击的是他们工作经历的早期,因为那时他们还没有能使自己对潜在的危险作出判断的情境意识。在危险情境中进行的研究常常取决于一种互动的"安全区"的发展。所谓安全区是一种围绕实地工作者扩展出来的短距离社会保留区,在这个区域里研究者感到安全,因而能得到心理上的放松。安全区的建立部分取决于那些在情境中的人对研究者的社会接受程度,部分取决于研究者估计进入和留在一个特定区域可能遇到的危险大小的能力。例如,一个和吸毒者打交道的研究者必须警觉与他们打架或爆发争端的可能性。此外,研究者还必须时刻小心飞车射击和警察的雷达探测系统。

个人的风格和行为举止在某些危险的情境中可能是很重要的。研究者可能会发现,必须要注意自己的衣着和行为,以免自己引起他人的过度注意,此外,他们也可能需要自信和得体的举止,以避免身处情境中的那些人感到他们正处于潜在受害者的位置。如果做不到这些,就有可能导致研究者受到攻击。在社会上受到某些知名人士和情境中人们信任的人"资助",有助于危险的避免。在潜在的风险情境中,诸如这样的个人可以警告研究者迫在眉睫的危险,或注意爆炸性的或有威胁的情境。有些研究者,例如那些工作在偏远地区的人类学家可能会遇到特殊的风险,即那些地区可能有传染病和寄生虫。在人类学家中,因落石和车祸致死或致残是很普遍的。

我们不应该认为在危险的环境中进行

研究是不可能的。事实上,涉及的那些风险有时可能是被夸大了。田野工作的危险是可以操控的,但这需要未雨绸缪,防患于未然。重要的是对危险的可能性和来源,以及研究者怎么做可能使危险得以被操控或加剧作出评估。以往对危险的处理过于随意,常把它们作为一种类似"战争故事的"资料,而现在,人们已经经常对田野调查中的风险进行讨论。对于员工和学生的保险和法律责任的担心促使我们对田野工作中的危险有了更清晰的了解。现在对田野工作危险的反映已经制度化,风险评估,预警指导,制订田野工作实施的政策框架,明确管理者的职责和人员培训等一系列措施已付诸实践。重要的问题在于,研究者不仅要意识到诸如这样的应对措施的发展,还要对此作出自己的贡献。

——Raymond M. Lee

(沈崇麟译校)

*也可参见伦理原则【Ethical Principles】、田野关系【Field Relations】、参与观察【Participant Observation】、研究的政治【Politics of Research】。

## 参考文献

Lee, R. M. (1995). *Dangerous fieldwork*. London: Sage.

# 数据(Data)

数据是正在研究的个案【Case】上记录的经验观察值,诸如这样的经验观察值包括国内生产总值的年度百分比变化、每个家庭的孩子数、态度量表上的得分、街头匪帮行为的实地记录或养老金领取者的访谈答案等。社会科学家从各种来源,包括实验、抽样调查、公共记录、历史文件、统计年鉴和直接实地观察收集数据。记录的经验观察值可以是定性的,也可以是定量的数据。使用定量的数据,经验观察值都是一些有着固定意义的数目,诸如以美元计的收入。而定性数据可以以文本形式或以已分配了数字编码的形式保存。不论什么方式可能都代表了一种可以对它们进行高效的记录和概括的方式。例如,在人类学研究中,研究者可以记录市场中的摊贩是用马车、牲口、汽车还是用自己的双手把货物运到市场。接着研究者可以用数字对这些观察进行编码,方法如下:马车 = 1,牲口 = 2,汽车 = 3,双手 = 4。这种编码易于将信息用计算机存储。计算机辅助的定性数据分析软件使用的上升势头很猛。最后读者应当注意英文中数据(data)一词是复数。

——Michael S. Lewis-Beck

(沈崇麟译校)

# 数据档案(Data Archives)

数据档案是获取、存储和传播研究及教学所需的二次分析研究数据【Data】的资源中心。其主要功能就是保证他们握有的数据能长期存储,并为将来所用。

数据归档是一种保存昂贵的资源并确保这些资源的研究潜能得到充分开发的方法。除非为将来的研究保存数据,并备有说明数据文件的文档,否则那些搜集费用昂贵、花费了调查研究人员和被调查对象大量精力和时间的调查数据到后来可能只留下为数不多的

几篇报告,只分析了数据潜能的一小部分。在很短的时间内,数据文件就有可能丢失,或者随着技术进步而变得过时了。

## 历史

1960 年代,数据归档运动开始在美国保存调查访问原始数据的许多重要的社会科学的教学单位展开。这场运动扩展到了欧洲,1967 年,(英国)全国社会科学研究会(National Research Council for the Social Sciences)建立了英国档案馆(UKDA)。1970 年代后期,许多国家档案馆都加入了更广泛的专业组织:欧洲社会科学数据存储委员会(the Council of European Social Science Data Archives,CESSDA)、国际社会科学数据组织联合会(the International Federation of Data Organizations,IFDO)。这些专业组织建立的目的是促进社会科学档案服务的网络建设,在关键归档策略、程序和技术上增进合作。

最初的数据档案收集的是社会科学中的定量研究【Quantitative Research】学者特别感兴趣的数据。1960 年代,这些数据大多是民意调查或选举数据,但随着大规模调查趋势逐渐呈现,到了 1970 年代末,英国国家档案馆开始获取主要的政府调查和人口普查数据。因为这些调查的样本规模巨大,收集的信息丰富,这些调查代表了主要的研究资源。英国政府主要的系列调查包括综合住户统计调查(the General Household Survey)、劳动力调查(the Labor Force Survey)、家庭支出调查(the Family Expenditure Survey)。美国政府主要的系列调查包括收入与(收入转移)项目参与研究(the Survey of Income and Program Participation)、当前人口调查(the Current Population Survey)以及国民健康访问调查(the National Health Interview Survey)。

到 1990 年代,英国国家档案馆馆藏的数据集已有数千种之多,涉及的数据极为广泛。以英国为基地的著名重复大型学术性调查包括英国社会态度调查(the British Social Attitudes Survey)、英国家户长期追踪调查(the British Household Panel Survey)和国家儿童发育研究(the National Child Development Study)。在美国则有美国的综合社会调查(the General Social Survey)、收入动态长期追踪研究(the Panel Study of Income Dynamics)以及 1972 年的全国纵贯调查。这些调查都是长期的系列调查的例子。主要的跨国系列调查包括世界价值观和欧洲价值观调查(the World Values and European Values Survey)、欧盟民意调查(the Eurobarometer Survey)、国际社会调查系列项目(the International Social Survey Program Series)。

1990 年代,研究界认识到了定性研究学者的需要,于是 1994 年在埃塞克斯(Essex)大学资助建立了定性数据档案库(Qualitative Data Archive,Qualidata),1995 年又成立了历史数据服务机构(History Data Service,HDS),致力于广阔的历史数据的归档保存和传播。

## 数据保有

社会科学数据档案得到的数据,范围极其广泛,涉及历史的和当今的社会,来源也是多种多样,包括调查、普查【Census】、登记(registers)和汇总统计数据。数据以各种各样的格式创建。数据档案一般都收集数值型的数据,这样的数据可以用统计软件来进行分析。将文字信息(例如对调查问题的回答)进行编码就产生了数字化档案,文本信息(如调查问题的答案)经过编码之后就变成了数值型数据,数值型数据也可以代表单个或汇总的数量,例如劳动收入。

访问数字文本、图像和视听材料的需求意味着来自深度访谈、田野笔记

【Fieldnotes】和录音访谈的材料以及开放式调查的问题现在都可以用于计算机分析处理了。

最后，通过国家数据档案馆与其他国家档案馆的互惠协议，整个世界的数据集都可以利用起来。

## 获取数据

数据档案关注的一个重要问题是，保证它获取的材料要适合于知情使用，并满足要求。所有保存的材料都经过了挑选和评估，都必须符合一定的标准，例如至少要有按最低标准准备的数据说明文件。不仅如此，获取档案的政策必须有灵活性，研究界能对数据和信息的需要变化以及技术环境的迅速变化做出及时的反应。

## 长期保存

数据档案的责任就是要关注硬件和软件发展，跟上技术进步的潮流，并对它们的藏品进行相应的变革。当技术发生变化时，档案馆要对其保有的数据做技术变换，以使它们在新环境下仍然可读。要将计算机程序维护成很容易就能将数据从机构内部的标准格式转换成用户需要的各种不同格式的状态。

## 准备需要归档的数据

获取数据时，数据档案保管员要进行数据处理工作。首先，要检查和验证数据集，例如要检查数值型数据的值，保证数据是匿名的，把与个人对上号的风险降到最低。其次，要生成“元数据”（有关数据的数据），其目的是生成高质量的档案检索工具和提供良好的用户文档。元数据包含描述研究和数据的信息，以及以后加到数值型数据文件的内容（如新建的变量和变量标签，以便用户查看变量信息和频次）。为方便研究，通常会创建一个系统的记载目录，记载研究的概况、数据集的大小和内容、数据集的可用度以及访问的条款和条件。用户指南进一步涵盖了如下信息：如何收集信息、原始问卷【Questionnaire】，以及如何使用这些数据。

## 提供数据访问

数据档案所提供的数据可以用于统计报告，例如基本的频数计数，也可以用于更深入的二次分析。二次分析加强了科学性的探究，避免重复收集数据，开拓了数据收集和测量的方法。重新利用档案数据可以让新的用户对旧数据提出新的问题；可以进行比较研究【Comparative Research】、复制【Replication】或再研究；可以对研究设计【Research Design】有所启发；还可以促进方法论的进步。最后，数据还可以为研究训练和实质性的学习提供重要的资源。

用户通常会要求获取某种特殊格式的数据，例如统计或文字处理包。时至今日，数据可以通过即时在线下载设备来访问，也可以通过光驱等便携的媒介来发送。21世纪出现了在线分析工具日益精细化的潮流，用户可以通过NESSTAR等网络浏览器来搜寻和分析数据子集。

数据档案提供的数据一般来说是匿名的，以便把与个人对上号的风险降至最低。通常用户需要签署一份协议，大致意思是说他们在进行分析时不会试图识别其中个人的身份。

——Louise Corti

（杨可译　沈崇麟校）

## 参考文献

ICPSR. (2002). *Guide to social science data preparation and archiving* [online]. Available: http://www.ifdo.org/archiving_distribution/datprep archiving bfr.htm.

IFDO. (n. d.). [online]. Available: http://www.ifdo.org/.

Ryssevik, J., & Musgrave, S. (2001). The social science dream machine: Resource discovery, analysis, and delivery on the Web. *Social Science Computing Review*, *19*(2), 163-174.

UK Data Archive (n. d.). [Online]. Available: http://www.data-archive.ac.uk/.

## 数据管理(Data Management)

我们需要以这样一种方式来管理研究数据:让它在不同年代和不同用户面前都能保持效用。这意味着我们必须关注数据的说明文件的编制存储和保存问题,这样至少其他人也可以访问我们的数据、复制我们的分析;若能对数据的效用有所扩展,则可谓锦上添花。我们如何才能实现这些目标,则取决于我们是生成这些数据(原始数据收集),还是仅仅再次利用数据(二次数据分析;参见二手数据【Secondary Data】,调查数据的二次分析【Secondary Analysis of Survey Data】),以及我们处理的是哪一种数据。

当我们讨论管理社会科学数据时,要区分两种最常见的数据类型:抽样调查【Survey】数据和文档数据。社会科学也处理其他类型的数据,例如实验【Experiment】数据和视听记录。音频或视频的记录会产生其自身的数据管理问题,可能广播这个行业的人对如何管理和归档这些数据有更深刻的见解,因此我们并不试图在这一领域提什么建议。经济学家越来越多地从实验环境中收集数据,以了解经济行为主体对被控制的经济机会会作何反应。这些实验需要用到用户自定义的软件,其本身便会生成用于分析的数据。总的来说,许多适用于调查数据的原则也适用于实验数据,但它的应用过于专业,因此,我们将不对这一问题进行过多的讨论。

原始数据搜集者承担着最为繁杂和重要的数据管理职责。为了能让其他人开发利用数据,收集者必须圆满地完成原始数据的管理任务。数据管理任务始于一个计划,该计划囊括整个研究项目,从数据收集到分析。这样一种数据收集的方式自然指明了数据管理的计划。对二手数据来说,管理数据则始于数据收集工作的下游,且数据管理问题的大小主要取决于数据收集者递交给用户的数据资源的种类。因而管理数据的最好策略可能是从原始数据收集者使用的数据系统开始。

对众多数据管理应用而言,关系型数据库技术主导着其他策略。大多数社会科学家都不熟悉这些系统是如何工作的,因此试图用他们熟悉的软件系统来管理数据,例如,SAS,SPSS 或 STATA。对小型项目来说,这可能是唯一可行的方法;学一门可能只用几个月的新技术也没什么意义,除非有人预见到自身的职业生涯会要求有很强的数据管理技巧。

我们从档案数据来开始我们对数据管理方案的讨论。这些数据特别适合于关系型数据库,而且采用档案数据也更容易说明这种技术的应用。尽管调查数据可能应用更广泛,但借助一个从行政案卷生成研究数据的例子,读者可能会更好地理解关系型数据的方法是如何被用于调查数据的。

### 行政案卷数据

档案数据有很多来源,但却几乎都来自研究者之外的某个人。档案数据采取何种数据管理结构才适当,取决于数据的复杂性。典型的档案数据都有一种简单的结构,并可以在用于分析的统计软件包内维护。

例如,国民收入账目和其他宏观经济数据都有一套固定的、在不同时期重复的变量。国民生产总值的数据是季度性的,失业率的数据是月度的,等等。股票交易的数据有各种依时加总的数据,乃至分解的全系列交易数据可资利用。

然而,有些档案数据可能是很复杂的。例如,美国各州都有海量的数据库,内容包括用于失业保险的收益报告、所得税,以及人们从诸如 TANF(有孩子贫困家庭前现金援助项目,始于 1990 年代末的福利制度改革,在各个州有各种特别的名称)、医疗补助计划(medicaid)和食物券(food stamps)等这样的转移项目的收益。这些数据包含了个人和家庭的档案,从申请、支付到行政裁定都包括在内。

我们以转移计划的行政数据为例,来描述一个关系型数据库策略是如何让研究者将数据组织为分析所用。为了阐述方便,我们将例子加以简化,所以读者不应该去推测会有任何一个特定的州用我们描述的方式来组织其数据。

关系型数据库包含了一套表格,每一个表都含有一套固定的变量。关系型数据库不要求表格以某种特殊的方式相关联,也不要求用某种特殊形式组合或提取数据。如果一个表格包含有关某个人(或某个公司、某项事务[transaction]等)的数据,我们就用"行"(row)这个词来指称那个表格中对应于特定的人(或公司、事项等)的一套变量。这种灵活性使得关系型数据库能很好地对来自各种不同应用的复杂数据,进行归档和检索。因为家户的规模和收入决定了福利性支出的规模,在福利数据库中的一个表格可能描述了一个接受某一特定项目资助的人。我们用援助群体(assistance group)这个词来表示有资格接受某个特定项目资助的那一群人。在一个单身户中,一个人可能既接受 TANF 项目的资助,又接受食物券项目的资助。因此,我们会因此假设数据库的数据描述了对应于某一收益类型的每一种援助群体,因而每种援助群体都有一个唯一的标识。一个 TANF 数据集一般会有一张标识受助者姓名、地址、人口特征(生日、民族、性别、国籍等)、社会保障号码,也可能会有表示每一援助群体的受益人的州识别数码。另一套表格记录有关签发给援助群体的支票具体事宜,用一条记录包括了援助群体标识、支付签发日期、支付的金额,以及任何与支付有关的具体的行政管理事宜的指标。同样,也可能会有一张存贮资格评估结果,或者社工家访记录的表格。

数据库中的其他表格可能也与个人有关。例如,对于每月需要资助的援助小组的人,在表格中会有一行用来记录那个人的姓名、出生日期、性别、民族、残疾状况、付款的项目,以及接受资助的援助群体的标识。

各州也维护着自己的数据库,其中失业保险项目包含季度收入。这个数据库里有一张表格含有工人的社会保障号码、企业的雇主识别号码、收入的季度,以及工人的收入数额。

用后面这套记录,我们可以建构出这个州的任何一个人的季度收入历史。同样,我们也可以给这个州的任何一个人构建一个逐月史,该历史显示该人在哪一个月得到了食物券或 TANF 的资助。我们也可以建立一个援助群体中的任何一个人的成员资格事件史,进而探讨它与工作收入之间的关系。我们可以用这些数据来确定,当某个小孩的父母失去福利收益时,小孩本人是否可以作为一个新的援助群体成员成为福利受益人。不同表格中同一个识别标志(如社会保障号码)的值相同的那些行可以连接起来。关系型数据库的优势并不在于有适合任何特定分析的正确的格式,而在于它使研究者得以将数据元素组合成适合任何实际分析的正确格式。而这也正是它能在商业应用中有明显优势之原因所在,从电子商务

网站到会计和库存系统,几乎无处不在。一些公司常常会在大型关系型数据库中的表格集合内整合它们所有的记录系统,提供非凡的追踪企业内部活动的能力。在一个设计精良的关系型数据库中,所有冗余的数据项目都会被删除。这会减少维护的费用和检索的错误。

当学者们明确地知道他们需要的是什么数据和它们必须以什么格式呈现时,他们可能就会从档案数据中抽取那些他们需要的变量,把数据转换成适合其研究方案的格式(例如,用他们的统计软件包可读的格式)。这种学者不太可能用到关系型的数据库,假设项目按预想的方向发展,这就是正确的决定。不过,这类研究者用原始数据所完成的工作可能对其他的数据使用者没什么价值,因为后者对他们所需要的数据和数据应如何组织另有自己的看法。这就违背了我们给数据管理设定的基本宗旨:保证数据在不同年代和不同用户面前都能保持效用。关系型数据库可能不是适用于任何特定应用的理想的数据管理系统,但它对几乎所有的应用来说都是非常好的系统。

## 调查数据

调查数据的结构常常很复杂,而且收集调查数据时通常都带有二次分析的想法,且对大多数有影响的调查而言,二次分析都是主要的目标。调查是关系型数据库技术最自然的应用,因为它常常是由一套套以多种方式互相关联的变量表格组成。

样本本身大多来自多级分层的【Stratified】随机样本【Samples】,而不是在教科书里看到的简单随机抽样【Simple Random Sampling】。此外,还通常都会访问从抽样框【Sampling Frame】中抽取的一个单位中的多个人。数据文件必须保存数据和抽样结构,以及被访者之间的关系。例如,在区域概率抽样【Probability Sampling】中,这意味着要保留每个观察在初级抽样单位和次级抽样单位中的标识,包括普查区和街区群的标识,以及抽样过程中每一步的选择概率。这些地理特点决定了各个不同的观察单位(被访者)相互关联的方式。近年来,地理信息系统(geographic information systems,GIS)数据的使用暴涨十分有说服力地指出,数据收集者应利用全球定位卫星接收器获得居住单位的经纬度,或使用Maptitude,ArcView等软件包给街道的经纬度做地理编码。经纬度数据是组织大多数GIS数据的基础,因此,这也应该成为位置的基本参数,而不是街区群、普查区或其他的索引系统。

调查数据通常会使用复杂的问卷,而这往往会给数据管理带来严重的问题。确实,随着计算机辅助技术在这个行业中占据主导地位,那些此前访问员不可能掌握的工具现在都成了通行的规则。我们鼓励对调查数据有兴趣的读者也去阅读一下计算机辅助个人访谈【Computer-Assisted Personal Interviewing】这一词条。该词条与本词条有很多相似之处,它与本词条一起,给读者提供了调查数据管理过程,从设计到数据传播的全貌。

工具常会收集一些人、雇主、保险计划、医疗服务提供者等的名单或花名册,然后再按这些清单循环提出一组有关花名册上列出的每个人、雇主或医疗服务提供者的问题。这些与调查问题有关的回答构成了一个更大的关系型数据库的一些表格,这些表格之间的连接则是由问卷设计所定义的。

我们可以把调查中的每个问题看成是具有多种属性的表格中的一行,这些属性以灵活的方式与其他表格相连接。调查问题的特点(或者“列”)大致如下:

- 问题的标识;
- 概括问题内容的描述语(饮酒、收入等);

• 问题的文本;

• 引入问题的一套问题或者检查项目;

• 一套许可的问题回答,以及这些许可回答的数据说明(无论答案是日期、时间、整数、币值、文本回答,还是分配给每一类回答的数值,例如 1=是,0=否);

• 导向下一个问题的路线指示,或者根据对当前问题的回答而定的检查项目;

• 帮助访问员或被调查者完成问卷的说明;

• 文字形式问题的备择语言版本;

• 对题项的准确性或解释,以及题项来源的说明。

我们常把问题的这些属性称为"元数据"。在复杂的调查中,这些描述问题的信息与问题是相关联的。例如,元数据包括哪些问题导向一个特别的问题,以及哪些问题是这个问题分支出来的。这些联系定义了问卷的控制流程或跳转模式。用一套复杂的描述问卷的表格定义,我们可以"合并"表格,进而快速地生成一系列报告,包括编码簿【Codebook】、问卷,以及其他各种传统的调查文件等。

如果我们把调查分解成一系列离散的事项(问题、检查项目、循环指令、数据存储命令等)并建立一个关系型数据库,每一个事项都是数据库表格中的一行,而表格有一套关系型数据库所定义的属性,我们就可以很有效地管理调查的内容和调查数据。

关系型数据库软件是软件产业的一个重要部门,诸如 Oracle, Sybase, IBM 和 Microsoft 等供应商都在提供具有竞争力的产品。许多商业应用使用关系型数据库系统(库存控制;会计系统;基于网络的零售、医院与福利机构的行政档案系统等),因此,社会科学家可以依托于一个成熟的软件市场。从关系型数据库的背景来看,建议提出的编码簿和调查数据文档的某些标准,例如数据文档的创立标准都和关系型数据库设计相似,但没有采用这些已有的成套的专业工具及它们的标准编程规范。

原始数据的收集者可以选择如下几种数据管理的办法:(a)围绕一个关系型数据库设计整个数据收集策略;(b)把后现场数据(post-field data)和工具信息输入一个关系型数据,或者(c)做一点其他的工作。对简单抽样调查来说,第三个选择可能就是用 SAS,SPSS 或 STATA 软件包来翻译编制调查数据文件。这些都是很优秀的统计软件包,我们可以用一个像 STATA 的 Stat/Transfer 这样的软件,在这些程序之间转移数据。这些统计软件包也开始融入了一些关系型数据库的特点。例如,SAS 支持 SQL(标准查询语言)对关系型数据库的查询,并已和关系型数据库关联起来。多年来,数据库管理一直在趋向关系型数据库方式,这一趋势的源头便是那些大型、复杂的项目。在设立那些大项目时,社会科学家最好要考虑确立自己的数据管理策略,并围绕关系型数据库管理系统来配备自己的工作团队。

——Randall J. Olsen

(杨可译　沈崇麟校)

* 也可参见定性数据管理【Qualitative Data Management】。

## 参考文献

Elmasri, R. A., & Navathe, S. B. (2001). *Fundamentals of database systems*. New York: Addison-Wesley.

Gray, J., & Reuter, A. (1992). *Transaction processing: Concepts and techniques*. San Francisco: Morgan Kaufmann.

Kroenke, D. M. (2001). *Database processing: Fundamentals, design and implementation*. Upper Saddle River, NJ: Prentice Hall.

Stern, J., Stackowiack, R., & Greenwald, R. (2001). *Oracle essentials: Oracle9i, Oracle8i and Oracle 8*. Sebastopol, CA: O'Reilly.

## 事后情况说明(Debriefing)

事后情况说明本来是一个军事术语,意思是在一个任务或一段服务期结束时进行询问或指导。在用于人的研究时,事后情况说明指的是在研究结束之后,调查者和被访人之间进行的交谈。事后情况说明是事前的知情同意【Informed Consent】的后期版,因此应本着一种有利于被访人和尊重被访人的方式来进行。

事后情况说明可能有几个目的。一般来说,对研究对象而言,这是一个提问的机会,而对于调查员来讲,这是感谢研究对象的参与、更充分地解释其研究和讨论研究对象对这一研究经历的看法的机会。参与研究可能对参加者有教育或治疗的价值,而事后情况说明是一个通过恰当的交谈和宣传巩固这些价值的恰当机会。事后情况说明还给研究者提供了一个让他们了解研究对象如何看待研究,以及为什么他们会如此行事的重要机会。感觉敏锐的研究者从事后情况说明中了解到的东西可能和研究本身一样多。为了实现事后情况说明的这些价值,最重要的是,研究者要准备好恰当的要传达的信息,带着敬意给研究对象提供一个提问和表达其想法的机会,营造一段感兴趣的研究对象和感激的调查员之间互动的轻松时光。事后情况说明不仅应该作为研究的不可分割的一部分来规划和安排,而且还应该和环境相适应。例如,如果研究的是私人或敏感的事,事后情况说明也应该在私下进行。假如研究的是儿童或行为能力受限的成年人,那么应该向研究对象的父母或监护人以及研究对象本人做出恰当的事后情况说明。

有时候,调查者承诺要给研究对象寄送研究的结果,而不是在对他们调查完之后和他们讨论这一研究。这样做有很多问题。这种承诺经常无法兑现。通常完成研究的时间常常会延迟很多,以致可能无法再与研究对象联系;如果能重新联系,信息到达之时,调查时的环境已不复存在,当时的经历也已遗忘殆尽。更具价值的是(本应)先于研究的从搜寻文献中获得知识,在事后情况说明时,可以很容易地用通俗的语言加以概括。

当研究涉及欺骗【Deception】(任何使研究对象认为有些事并非如此的程序)时,事后情况说明中也应当包括对欺骗性质的说明及之所以这样做的原因(欺瞒情况报告,dehoaxing)。在说明了欺骗或紧张情况之后,事后情况说明还应该报告脱敏(desensitizing)情况,即为让研究对象回归积极的情绪状态做过的一些工作。如果导语诚恳、直接地让研究对象同意参与各种特定的指定条件中的一种(例如在安慰剂研究中)同意被欺骗或放弃知情权(对这些方法的讨论参见 Sieber, 1992)的话,欺骗问题也很容易在事后情况说明中报告出来。

这样的研究对象期望事后情况说明能揭示真实情况。相反,那些得到了错误信息的研究对象可能会在说明真相之后有负面反应,感觉他们被要弄了,他们可能会怀疑自己是否可以相信欺瞒情况报告。在这种不信任的情况下进行脱敏大概是不可能的。同样,那些不知道他们参加了研究的研究对象,可能会坚持相信他们参与的是一个自然发生的(不受控制的)情景,可能较难进行事后情况说明。不管用什么手段进行欺骗(例如,一个同谋,对某种东西或某个程序的虚假陈述),一般多不难安排有说服力的欺骗示范。例如,如果之前给研究对象的是一个测试成绩的虚假报告,欺瞒情况报告则可包括把研究对象自己的测试成绩给他们,就像他们当初上交时一样,仍然密封在一个信封里。

## 什么时候不需要做欺瞒情况报告

欺瞒情况报告可能涉及包含对研究方案和程序的解释，包括所做的假设和比较。不过，事后情况说明应该适合研究对象的文化和教养程度，其内容不应涉及对研究对象的责备、贬损或迷惑。因此，精细的说明可能并不合适，例如，在下列情况中，研究对象很可能会不理解详细的欺瞒情况报告，或从报告中得到什么好处：

- 对儿童忍住不拿（偷）不属于他们的东西的能力研究，并将这一能力作为道德发展阶段的一个函数；
- 将支持体罚的父母群体和支持其他手段的父母群体进行比较的对于子女抚育习惯和儿童攻击性的研究；
- 对常用恶灵或巫术解释问题的文化成员进行的疾病解释研究。

提供一个简单的、没有偏见的研究描述可能不失为一种明智之举，例如，"我们想了解父母用来惩戒孩子的各种不同方法。我们非常感谢您让我们观察您和您孩子的互动。"如果研究对象希望了解多一点，那么我们可以提供一些其他的无偏见的信息，例如，"这是我们观察到的几种方法"。

设计适当的事后情况说明需要对研究对象的需要、文化和感情很敏感。事后情况说明的主要目的就是感谢研究对象的参与；回答他们的问题；了解他们的反应；以及用一种有礼貌的方式，把可靠的，也许有所教益的信息传达给研究对象。

——Joan E. Sieber

（杨可译　沈崇麟校）

## 参考文献


American Psychological Association. (1992). Ethical principles of psychologists and code of conduct. *American Psychologist*, *47*, 1597-1611.

Holmes, D. (1976). Debriefing after psychological experiments: Effectiveness of post-experimental desensitizing. *American Psychologist*, *32*, 868-875.

Sieber, J. E. (1992). *Planning ethically responsible research.* Newbury Park, CA: Sage.


# 欺骗(Deception)

欺骗是在告知被调查者研究正在进行，以及研究的性质、目的和后果时，采用一种失当或蓄意误传的做法。

在社会研究中某种欺骗测量是广泛存在的，尽管这些并没有全部被认定为欺骗。研究者告知参与者以获取他们同意参与的故事往往是片面的。因为研究者害怕一旦他们揭示了全部事实，他们会失去自愿的被调查者。他们还会对他们的研究兴趣使用误导性描述，例如，当他们意在性和毒品问题时，却去谈论什么娱乐消遣和健康问题。

然而在 1960 年代和 1970 年代，社会研究者还没有发展出集体良知以前，欺骗更是不加掩饰的。研究者进入研究领域时通常都采用角色扮演、伪装或伪造身份的形式。例如，罗森汉恩（参见 Bulmer, 1982）及其合作者模仿精神病患者的症状，以获取精神病机构和试验甄别程序的许可。哈沃德・格里芬通过服用药物改变了自己的肤色，以实施一项有关美国南方腹地黑人的伪参与研究。出乎意料的是，这个程序是不可逆的。肯・普赖斯为了获得布里斯托尔一个加勒比黑人教会的牧师的合作，他表示相信自己并不相信的信仰，并接受了洗礼（Homan, 1992）。

斯坦利·米尔格拉关于服从和权威的臭名昭著的研究,则采取了另一种形式的欺骗。米尔格拉误导那些被雇用的参与者,使他们认为自己是米尔格拉的研究助手。然后米尔格拉诱骗他们相信自己正在造成其他的研究参与者遭受疼痛,而这些另外的研究参与者实际上都是受过训练的演员。因为研究参与者最终愿意引起他人的疼痛和悲苦,所以他们后来都会因此遭受各种不利的后遗症。

米尔格拉和其他一些研究案例引发了人们对于欺骗对研究参与者所产生影响的关注,同时也引发人们关注研究专业人员的名声。在个体层面,欺诈和背叛可以变成一种习惯行为和一种不信任感。对于研究者的共同体而言,存在一种他们以不光明正大和欺诈的方式获取名誉的风险。对于一个接受社会科学训练而非戏剧学校出身的学生而言,扮演一个不熟悉的角色会为他们带来不小的压力。改变肤色、入住精神病院或加入犯罪团伙并不是很多人想要承担的任务,更不要说让自己的学生承担。

但是,社会科学中的欺骗在道德上是否总是错误的,仍然是一个有待讨论的问题。我们同情那些因参与服从实验(obedience experiments)而遭受精神痛苦的参与者。我们谴责那些实施了这些项目的人,并认为对欺骗的普遍禁止将会预先阻止这些不利后果。但当欺骗用来避开强势群体或社会中受谴责的群体(如三K党)时,对欺骗的愤慨之情就会缓和许多。对于一些能够照管自身,但却拒绝接受常规调查的群体而言,可能有一定理由来使用欺骗策略。研究者必须考虑到,他们对他们的研究对象所负有的责任和对真实情况予以公开的责任之间是有矛盾的。

——Roger Homan

(郑晓娟译　沈崇麟校)

## 参考文献

Bulmer, M. (Ed.). (1982). *Social research ethics.* New York: Holmes and Meier.

Homan, R. (1992). *The ethics of social research.* London: Longman.

Lauder, M. (2003). Covert participant observation of a deviant community: Justifying the use of deception. *Journal of Contemporary Religion*, *18* (2), 185-196.

Milgram, S. (1974). *Obedience to authority.* London: Tavistock.

# 十分位间距(Decile Range)

十分位间距的计算方法和作用与四分位距【Interquartile Range】类似。先将一个频数分布从高到低排列,再将它分成十个,每个组有相同的个案数的组。将第一个和最后一个十分位去掉,而十分位间距就是那些剩下的值中的最低和最高值之间的差。与四分位间距一样,十分位间距也是一种处理影响间距【Range】的特异值【Outliers】问题的方法。这一词,有时也作为"十分位数"的同义词使用,不过这种用法是不够正确和精确的。

——Alan Bryman

(沈崇麟译校)

## 解构主义(Deconstructionism)

解构主义主要是将那些生产现实的特定版本的范畴和策略加以确认和问题化。它对文本进行剖析,以观察它们为什么以这样一种方式建构,来展现人、地方和物体的特定形象。作为后解构主义运动的一部分,法国社会理论家雅克·德里达(Derrida,1976)首先在1960年代末将解构主义引进哲学。尽管解构主义和文学理论关系紧密,但也在建筑、艺术、文化、法律、媒体、政治和心理研究中引起了不少反响。

解构主义立足于对文本的深入阅读,使我们能确定文本建构是如何依赖于矛盾、困惑和未言明的缺失的。德里达反对主导西方思想和哲学的那些固定的范畴逻辑,尤其是二元对立非此即彼的逻辑,例如,灵/肉、男/女、健康/疾病、个人/社会。德里达提出,我们要接受既/且(both/and)逻辑,而不是二元对立的非此即彼(either/or)逻辑。当现象被认为似乎是既包括又排斥时,它们才能更好地得到理解。例如,个人/社会的对立或者可以理解为一个系统的结合特性,每个组成部分都需要他方的存在,并赋予其存在的意义(Burr, 1995)。对立在一个文本中的作用是,充当突出伦理、意识形态和政治过程的手段。

伊恩·帕克(Parker, 1988)建议采用三个步骤或阶段来进行解构主义的分析。第一步是确定文本中的二元【Binary】对立或两极,并对两极中有一个特定的极处于优势地位的对立方式加以说明。第二步则把对立中处于从属位置的一极转变为支配位置。这样就能说明优势的一端依靠的是什么,以及为什么没有了对立的另一极,它就无法存在,因而也就不能起作用。第三步则对对立予以重新考虑,进而发展出新的概念和实践。帕克也用了这种解构主义的三阶段来分析英国广播肥皂剧"阿基一家"(The Archers)的话语。通过对这部肥皂剧的转录文本的分析,帕克对在讨论性别问题时使用的现实和虚构、日常解释和理论解释之间的区别提出了抗争。

解构主义是有争议的。德里达被誉为对当代思想做出了最重要的贡献者之一,同时也因破坏了广为接受的文本、意义和身份而受人诟病。支持解构主义的人们认为它提供了一种研究语言和意义的新奇而有趣的方式,它可以破坏理论,并开启讨论与竞争之门。同样地,解构主义也可以用于变革权力关系,鼓励推翻主导性的思想观念。不过,反对者们声称解构主义关注的是验证那些无法验证或证明的东西。通过解构主义来证实特别困难,因为其质询的对象是不稳定的。我们无法确定一个文本的意义,因为它是由一串没有穷尽的意义链组成的。因此,解构主义的分析仅限于举例说明一个文本中的语义。解构主义受到批评的另一个原因在于,它在理解语义建构和权力斗争时过于注重文本的作用,而对受众作用的考虑则有所欠缺。

——Sharon Lockyer

(杨可译　沈崇麟校)

### 参考文献

Burr, V. (1995). *An introduction to social constructionism*. London: Routledge.

Derrida, J. (1976). *Of grammatology* (G. Spivak, Trans.). Baltimore, MD: Johns Hopkins University Press.

Parker, I. (1988). Deconstructing accounts. In C. Antaki (Ed.), *Analysing everyday explanation: A casebook of methods* (pp. 184-198). London: Sage.

## 演绎(Deduction)

从逻辑上说,演绎是用来从一般性的陈述中推导特殊陈述的过程。社会科学中用这个逻辑来检验理论【Theory】。假设【Hypothesis】是从理论中推导出来的,并通过比较相关的数据【Data】来进行检验。如果数据和假设不符,那么这个理论就被视为错误理论而否定;如果数据和假设相符,理论就暂时被接受,从而得到了证实。检验这样一个假设被当作对理论本身的检验。

演绎的概念很古老,可以追溯到欧几里得几何和亚里士多德逻辑学。演绎是假说-演绎法【Hypothetico-Deductive Method】本质的一面和卡尔·波普尔的证伪主义【Falsificationism】有关联。波普尔主张采用演绎的逻辑来克服归纳【Induction】的缺点。他主要的观点是,因为观察数据不能为科学理论提供可靠的基础或是归纳的逻辑有缺陷,需要采用一个不同的逻辑来发展理论。他的解决方法是承认所有的数据采集都是有选择性的,而且需要观察者的诠释,那么就要采用一种与归纳相反的逻辑。

波普尔主张科学的目的就是要试图去否定已提出的试验性理论,而不是像归纳法那样,去寻找支持正在形成的普遍性的确定证据。真理的寻找是难以捉摸的,因为我们无法知道我们何时能到达。我们能做的全部事情,就是通过显示错误理论和数据不符来对它加以排除。通过了检验的理论并没有被证明是对的,因为进一步的检验还可能会发现理论不成立的证据。可能必须放弃一种理论,或至少要对它进行修正,做进一步的研究。因此,科学是一个推测和拒绝的过程、一个试错的过程、一个建造试验性的理论,并进而努力证明该理论错误的过程。

演绎性理论的例子可以参见涂尔干对利己型自杀的重构。他采用了三个概念,构建了五个命题。这三个概念分别是:自杀率(某个人口群体中每千人自杀数)、个人主义(人们不服从某个群体的信仰和规范、自行思考和独立行动的倾向)以及新教(在宗教改革和罗马天主教分裂之后兴起的一些基督教群体)。他的五个命题如下:

①在任何一个社会团体中,自杀率随个人主义(利己主义)程度的变化而变化。

②个人主义程度直接随新教徒的比例变化。

③因此,自杀率随新教徒的比例变化。

④西班牙新教徒比例低。

⑤因此,西班牙的自杀率低。(Homans, 1964:951)

这番论证的结论就成了有待检验的假设。如果假设被否定,这个理论就被认为是错误的。例如,改换国家,不仅可以做预测,而且假设也可在其他背景中检验。

如同归纳逻辑一样,演绎逻辑也遭到了激烈的批评。主要的批评包括如果观察都是诠释,那么我们就永远也无法直接观察到真实,无法自信地建立规则,也无法最终拒斥某些理论;试验性地接纳尚未被拒斥的理论,需要归纳法的支持;太过重视逻辑可能扼杀科学的创造力;接受或者拒绝某些理论的过程涉及社会和心理过程,而不只是逻辑过程

（评论请参见 Blaikie，1993；Chalmers，1982）。

——Norman Blaikie

（杨可译　沈崇麟校）

## 参考文献


Blaikie，N.（1993）. *Approaches to social enquiry*. Cambridge，UK：Polity.

Chalmers，A. F.（1982）. *What is this thing called science?* St. Lucia：University of Queensland Press.

Homans，G. C.（1964）. Contemporary theory in sociology. In R. E. L. Faris（Ed.），*Handbook of modern sociology*（pp. 951-977）. Chicago：Rand McNally.

Popper，K. R.（1959）. *The logic of scientific discovery*. London：Hutchinson.


# 自由度（Degrees of Freedom）

参数估计【Parameter Estimation】可资利用的独立的观察数即所谓的自由度（*df*）。一般来讲，一个检验统计量的自由度取决于数学上施加于被估计的数量的约束。例如，自由度必须与 *t* 检验【*t*-Test】、*F* 比率【*F* Ratio】、卡方【Chi-Square】、标准差【Standard Deviation】或回归【Regression】分析一起计算。在回归分析中，自由度的计算方法是样本量减去包括常数项在内的系数数目。

考虑一下用一个样本量为 *N* 的随机样本估计总体标准差【Standard Deviation】的简单例子。其计算公式为：

$$SD = \sqrt{\frac{\sum (U - \bar{U})^2}{N - 1}}$$

式中的 *SD* 为估计的样本标准差，$\sqrt{\ }$ 为平方根，$\sum$ 为和，*U* 为总体变量上的样本观察数，$\bar{U}$ 为样本观察的平均分，而 *N* 为样本量。注意，分子除以了 $N-1$，而不是 *N*。实际上这个修正似乎很小，但在理论上却是很重要的。为了避免总体标准差的估计值的夸大和偏倚【Bias】，这个修正是必需的。在计算这个总体参数时，我们耗费掉了 1 个自由度。这是因为在公式中使用了样本均值这一特征值。共有 *N* 个 $U-\bar{U}$ 的分数，但只有 $N-1$ 个是独立的，因为一旦我们要计算它们，这最后一个，第 *N* 个的值就必须固定。这一点来自总是有 $\sum (U-\bar{U}) = 0$ 这一事实。例如，假设 $N=3$，$(U_1-\bar{U})=6$，$(U_2-\bar{U})=12$。那么就必定是 $(U_3-\bar{U})=-18$。对于偏离均值的约束，它们的和为零这一点表明最后一个偏差并非“自由”的，而是固定的，所以丢失了 1 个自由度。在这个例子中，我们要估计的总体参数只有 1 个，即标准差，故而剩下的自由度为 $N-1$ 个。

在更为复杂的回归分析中，$df=N-K$，式中的 *N* 为样本量，而 *K* 为估计的参数数。假设多元回归模型为 $Y=a+bX+cZ+e$，$N=45$。要估计的参数有三个，即截距 *a* 和两个斜率 *b* 和 *c*。因此，$df=N-K=45-3=42$。在进行回归分析时，要有足够的自由度是很重要的。举一个极端的例子，在简单相关【Simple Correlation】（回归【Regression】）模型 $Y=a+bX+e$ 与 $N=2$ 的样本拟合时，$R^2=1.0$，因为普通最小二乘【Ordinary Least

Squares】拟合的直线连接两个数据点是无误差的。这并非是一个完美的解释，而仅仅只是数学的必然性而已，因为自由度已经丧失殆尽，也就是说，$N-K=2-2=0$。一般来讲，在回归分析中，较之独立的变量，我们希望有更多的独立观察，而不是更多的变量。在 $K$ 开始接近 $N$ 时，分析师将对调整的 $R$ 方【Adjusted $R$-Squared】予以特别的关注，因为它会对丢失的自由度进行修正。

——Michael S. Lewis-Beck

（沈崇麟译校）

### 参考文献

Kennedy, P. (1998). *A guide to econometrics* (4th ed.). Cambridge: MIT Press.

Lewis-Beck, M. S. (1995). *Data analysis: An introduction*. Thousand Oaks, CA: Sage.

## 删除(Deletion)

删除是一种处理缺失数据【Missing Data】的方法，缺失数据通常都是在开始分析之前被排除（转换）的。计算机删除缺失数据最常见的类型被称为整条删除【Listwise Deletion】缺失值（也被称为整个个案删除，casewise deletion）。在采用整条删除时，只要分析的任何一个变量上有缺失值，这个个案就要全部删除。整条删除的一个替代方法是成对删除【Pairwise Deletion】。在采用这种方法时，只有在一对研究变量上都有缺失值时，这个个案才会被全部删除。在删除之后，我们就可以正式开始进行分析了，例如，从估计相关【Correlation】矩阵开始进行分析。成对删除策略的优点是可以让样本【Sample】保持一定的容量，而缺点则在于可能会引起结果的统计不一致性。

——Michael S. Lewis-Beck

（杨可译　沈崇麟校）

## 德尔菲法(Delphi Technique)

这是一种在一组专家中取得一致的定性方法。其名字源于古代善于预言未来的德尔菲神谕。兰德公司（RAND Corporation）1960年代发明了这种办法，用于预测未来事件发生的可能性、合意性及其影响。现在德尔菲法的应用已经十分广泛，凡是希望意见集中的地方，几乎都可以使用德尔菲法。现在有三种不同的形式：探索性德尔菲法（exploratory Delphi），这也正是兰德公司开发出来的方法；焦点德尔菲法（focus Delphi），用于征求可能会受到某一政策影响的不同团体的意见；规范德尔菲法（normative Delphi），用于收集各个专家对某一定义明确的议题的观点，达到共识（例如，制订宗旨和目标）。德尔菲法过程的基本要点是，参与者在发表意见时都需匿名，这样可以减轻整个群组被少数几个强人控制而产生的问题，也可由此得到定量的估计（例如，开支优先顺序以及相应的款项分配）。

德尔菲法是一个如下文这样的迭代过程：

第一轮。参与者受邀对工作组(facilitating group)所提供的一个主题或一组观点进行评价。这些观点应该以参与者的个人知识和经验为基础。接下来工作人员就分析结果提供反馈，例如，图形概括、排序或者在有限几个标题下列表。接下来，将这些结果作为第二轮的问卷反馈给参与者。

第二轮。参与者在第一轮反馈的基础上对其他人的观点摘要进行评论，也重新审视自己的观点。工作人员和第一轮一样，对这些观点加以分析，再把结果作为下一轮的问卷反馈给参与者。这种轮次重复的过程一直继续下去，直到意见达至"充分的"收敛。通常三轮就足够了。参与者除了要记下他们对结果的同意程度之外，常常还要回答他们对自己的观点的置信度——如果较多的专家参与者的观点并非那么适中的话，这一置信度还可以用来给答案加权，以产生概括性的结果。随着互联网的进步，以及将来网格(The Grid)时代的到来，现在正是我们把工作人员的许多分析加以自动化，并加快这个过程的大好时机。

非常重要的是，参与者都是在某种意义上的"专家"，这样才能得到有价值的结果，而且应该有足够多的参与者，这样才能避免少数人的极端意见不会对整个群组的结果产生不成比例的影响。对某种服务的使用者而言，他们可以被视为"需要服务提供什么"这个问题的"专家"，但是就潜在的风险和好处来说，专业人士可能更了解情况。二者可以处于同一个研究之中，同时要求每个组都要考虑其他组的观点(Charlton, Patrick, Matthews, & West, 1981)。还需要在一开始就考虑在每轮之后要给参与者提供什么反馈，以及"同意"如何来定义，还有如何处理游离值。在研究结束时，还需要考虑结果的效度和可接受度。德尔菲法最为适用的场合是寻求各种意见，而非用常规的统计方法综合证据的时候。德尔菲法可能更应被视为一种组织群体沟通的方法，而非提供确定性回答的方法。

——John R. H. Charlton

(杨可译　沈崇麟校)

## 参考文献

Charlton, J., Patrick, D. L., Matthews, G., & West, P. A. (1981). Spending priorities in Kent: A Delphi study. *Journal of Epidemiology and Community Health*, *35*, 288-292.

Linstone, H. A. (1978). The Delphi technique In R. B. Fowles (Ed.), *Handbook of futures research.* Westport, CT: Greenwood.

Jones, J. M. G., & Hunter, D. (1995). Consensus methods for medical and health services research. *British Medical Journal*, *311*, 376-380.

Murphy, M. K., Black, N. A., Lamping, D. L., McKee, C. M., Sanderson, C. F. B., Ashkam, J., & Marteau, T. (1998). Consensus development methods, and their use in clinical guideline development [entire issue]. *Health Technology Assessment*, 2(3).

Pope, C., & Mayes, N. (1995). Qualitative research: Reaching the parts other methods cannot reach: An introduction to qualitative methods in health and health services research. *British Medical Journal*, *311*, 42-45.

## 需求特征(Demand Characteristics)

在一些研究中,非常典型的是在基于实验室的实验性研究中,研究者的期望是非常明显的,研究假设易于识别,研究者期望获得的反馈也是显而易见的。我们说这样的研究情境便是具有高度实验需求特征的。需求特征是(研究者、研究情境等)提供给研究参与者的暗示,这种提示虽然常常是无意识的,但却让参与者知道怎样的表现是合适或者被期待的。当参与者给出了研究者正好想要的回答时,研究者的一个微笑,一个轻微的点头,一个声音语调的细微变化——这些行为中的任何一个都可能对调查结果造成偏差。需求特征产生了由对实验过程的不当操作而归因出的结果,因而它严重损害了研究的内在效度【Internal Validity】和外在效度【External Validity】。当结果的变化是由需求效应引起时,涉及将这种实验处理认作观察结果的原因的解释是错误的。

罗伯特·罗森塔尔(Rosenthal, 1976)最为坚定地致力于引起社会科学家对实验需求的曲解效应的关注,在一系列研究中,他揭示了研究者的期望是如何被转移到研究参与者的身上的。在他最有争议的研究中,有一项指出了教师对学生未来成就的期望影响了学生的实际成就获得(Rosenthal & Jacobson, 1968),后续的研究分析证实了他的解释。基于对需求特征的潜在偏误效应的高度关注,现在社会科学家已经非常注意尽量隐藏研究假设,并经常借助精心设计的表面故事【Cover Story】来隐藏研究的真实目的。除此之外,为了避免对于所期待的反馈的无意暗示,研究经常采取一种无论研究者还是参与者都不清楚实验特定条件,且个体都是(随机)分配的研究方式。目前这样的双盲【Double-Blind】技术在社会科学中十分常见,这代表了对于需求特征可能破坏研究结果的强有力抵制。

需求特征与研究参与者采用的特定心理定式有关,也与他们对于自身作为研究被试者的自我意识有关。科拉诺和布鲁尔(Crano & Brewer, 2002)将研究参与者分为三个类别:自愿的研究参与者,他们对于自己作为研究被试的角色有清楚的认识,并对此感到满意;不自愿的参与者,他们认为自己是受到强制来服务于研究的;非自愿的参与者,他们没有意识到自己在参与研究调查。在这三种类型中,自愿的研究参与者最有可能受到需求特征的影响,因为他们是自觉作出参与决定的,他们也没有感觉受到不恰当的强制胁迫;因此,他们可能更易于试图通过自己的表现,确认研究中明显呈现的假设来"帮助"研究者。不自愿的研究者由于感觉到自己是被强制服务于研究的,因而更有可能暗中破坏研究。非自愿的研究者不太可能试图帮助或者破坏研究,因为从定义来看,他们还没有意识到自己正在参与研究。

——William D. Crano

(郑晓娟译　沈崇麟校)

### 参考文献

Crano, W. D., & Brewer, M. B. (2002). *Principles and methods of social research* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum.

Rosenthal, R. (1976). *Experimenter effects in behavioral research*. New York: Irvington.

Rosenthal, R., & Jacobson, L. (1968). *Pygmalion in the classroom*. New York: Holt, Rinehart and Winston.

## 民主研究和评估(Democratic Research and Evaluation)

慎重的民主研究和评估力求考虑各相关方的利益、价值观和观点;通过主要的利益相关者的广泛对话和对研究结果的进一步的深思熟虑来得到无偏的结论。

——Ernest R. House

(郑晓娟译　沈崇麟校)

### 参考文献

House, E. R., & Howe, K. R. (1999). *Values in evaluation and social research*. Thousand Oaks, CA: Sage.

## 人口学方法(Demographic Methods)

人口学方法包括各种不同的用于描述人口【Populations】的量度和技术。虽然它们一般都应用于人类的总体,但是现在也已经越来越多地用于描述非人类的各种总体,如禽兽、昆虫和工业产品等各种总体。

最基本的人口学技术,如比例和比率的计算,描述了总体的组成,如年龄-性别结构、种族组成、地理分布。诸如比率和概率这样的量度则描述了人口的动态。某些人口学方法专注于产生人口变化的单独过程。例如,单个和多个衰减生命表【Life Tables】,专注于退出过程,如死亡率。比较复杂的技术则同时考虑进入(如生育率)和退出二者,将它们结合起来估计生殖力、内在的增长或下降率及平稳和稳定人口的参数等。诸如这样的直接标准化、份额计算(calculation of shares)或相异指数都允许进行跨两个或更多个总体的比较。

### 数据来源

用人口学方法描述人口总体一般都以人口普查、抽样调查、人口动态登记系统,有时也以临时人口登记产生的数据为依据。人口普查【Census】的目的是得到一个特定日子的完整的所有人口计数,通常是在一个国家现有的所有人口计数(实际计数)或所有居住在一个国家的所有人口计数(法律上的计数)这两种计数中的一种。实际上,普查计数会在某种程度上,受到某些个人的统计漏查或少计(常通过捕捉—再捕捉法【Capture-Recapture】模型加以确认),或其他个人的多计的影响。大多数国家的人口普查也包括了每个普查计数的个人的有关信息,如年龄、性别、与户主关系、教育程度和婚姻状况等。

人口动态登记系统描述出生、死亡,有时也涉及其他的人口学事件(如结婚和离婚),因为它们都随时间的推移发生。人口动态登记系统经常也包括一些其他信息,如年龄、性别和婚姻状况等。人口登记虽然通过不断地追踪整个人口和极其重要的事件,将定期的人口普查和动态人口统计系统二者的长处融合在了一起,但是它们不仅比较难以操作,而且费用也比较昂贵,因此很少使用。

### 基本的人口学量度

比例(或百分比)和比率可以用来描述人口的结构,或它的成员沿一个感兴趣的变量的分布。因为许多人口学和社会过程在

人口中的变化是有着非常强的年龄等级的，比较复杂的人口学技术常常会将年龄结构列入控制或考虑之列。人口完整的年龄结构可以通过将绝对的人口数列表，或计算每一年龄间隔的比例(或百分比)来描述。得到的结果常常被绘制成一张人口金字塔图。我们也经常通过计算一些概括性量度来描述人口年龄-性别组成的特定性质。一个重要的概括性量度是年龄抚养比，它是以年龄在 18 岁以下和在 65 岁及以上的人口数("依靠"人口)，除以年龄在 18~64 岁的人口数("工作"人口)来定义的。另一个概括性量度是性别比，它是以男性人口数除以女性人口数来定义的。

## 变化的量度

几乎所有描述变化的人口学方法和技术都建立在一个基本的人口统计学模型上。在那个模型中，在时间 $t$ 的人口量是在较早的一个时间点 $t-n$ 的人口量及时间 $t$ 和时间 $t-n$ 之间进入和退出数目的函数。例如，国家人口在时间 $t$ 和时间 $t+n$ 之间的变化，体现在四个经典的人口学事件：出生($B$)、死亡($D$)、迁入人口到达($I$)和迁出人口离开($O$)上。

最基本的描述变化的量度是比率。比率是发生在一个特定的时间范围内(一般是 1 年)的人口学事件数，与有经历过此事件的风险的人口数之比。例如，人口粗死亡率，可以用在一特定年死亡的总人数，除以今年的总人口数的估计值来计算。比例常会乘以一个常数 $k$，如 1 000 或 10 000。例如，2000 年美国粗死亡率(CDR)，为每 1 000 人中有 8.5 人死亡，其计算公式为

$$\text{CDR} = \frac{D}{N} \times k = \frac{2\ 403\ 000}{281\ 422\ 000} \times 1\ 000 = 8.5$$

美国 2000 年的粗出生率(CBR)以与上面类似的方式定义(只不过用出生数代替了死亡数)，为每 1 000 人中有 14.4 人出生。

$$\text{CBR} = \frac{B}{N} \times k = \frac{4\ 065\ 000}{281\ 422\ 000} \times 1\ 000 = 14.4$$

CBR 和 CDR 的差就是粗自然增长率(CRNI)。在美国，2000 年的年自然增长率是每 1 000 人有 5.9 人。

$$\text{CRNI} = \text{CBR} - \text{CDR} = 14.4 - 8.5 = 5.9$$

粗迁入率或粗迁出率可以直接以类似于粗出生率或粗死亡率的方式计算。但是许多国家都没有收集全面的迁出的数据，而有关迁入的数据往往都是有瑕疵的。因此，迁移率往往都指的是净迁移($I-O$)。

## 死亡率

比较精确的比率涉及用年龄、性别或社会特性，如族群或种族定义的子总体。因为死亡的力量(the force of mortality)随年龄的变化十分显著，因此死亡率常常分年龄组列出。下面给出的分年龄组的死亡率使用了常用的记号方式，在这样的方式中，大写字母 $M$，$D$ 和 $N$ 分别指死亡率、死亡数和人数。而右边和左边的下标则分别确定了下限和年龄组的宽度。2000 年，美国年龄为 15~24 岁的人的死亡率是每 100 000名年轻成年人中有 79.9 人死亡。

$$_{10}M_{15} = \frac{_{10}D_{15}}{_{10}N_{15}} \times k = \frac{31\ 300}{39\ 183\ 900} \times 100\ 000 = 79.9$$

如果我们关注的问题是婴儿，那么常常将年龄为 0~1 岁人口中的婴儿死亡率(IMR)来替代分年龄的死亡率，因为人口普查中婴儿的计数常常有误差。婴儿死亡率的计算方法是将死亡数除以出生数。2000 年，美国从有关婴儿死亡和出生的初始数据

得到的婴儿死亡率是每 1 000 名出生的婴儿中有 6.9 人死亡。

## 生育率

生育率一般只涉及处在生育年龄的那些妇女,因为难以收集到有关父亲年龄(或其他属性)的数据。在下面计算分年龄组的生育率的公式中,大写字母 $F$,$B$ 和 $N$ 分别指生育率、该年龄组妇女的生育数和该年龄组的妇女数。与前面的公式一样,下标表示年龄组的下限和组宽。2000 年,美国 15~19 岁妇女的生育率为每 1 000 名妇女生育 48.5 人。

$$ {}_5F_{15} = \frac{{}_5B_{15}}{{}_5N_{15}} \times k = \frac{469\ 000}{9\ 670\ 000} \times 1\ 000 = 48.5 $$

将育龄期(一般为 15~44 岁或 15~49 岁)妇女的分年龄组的生育率加起来,并乘以年龄组的宽度,便可得到总生育率(TFR)。列在表 1 第一列的分年龄组的生育率之和显示,2000 年,美国的总生育率是每 1 000 名妇女生育 2 130 人。TFR 可解释为在 2000 年观察到的经历了特定年龄生育率的假设的妇女同期群,在她们活着经历了自己的生育期后的生育总数。

**表 1　2000 年,美国人口再生产情况一览表**

| 年龄组 | 分年龄组每 1 000 名妇女生育率 | 分年龄组出生率×女婴出生比例 | 活到母亲的年龄的女儿的比例 | 乘积 |
|---|---|---|---|---|
| 15~19 | 49.4 | 24.1 | 0.990 2 | 23.9 |
| 20~24 | 112.3 | 54.8 | 0.987 9 | 54.1 |
| 25~29 | 121.4 | 59.2 | 0.985 4 | 58.4 |
| 30~34 | 94.1 | 45.9 | 0.982 2 | 45.1 |
| 35~39 | 40.4 | 19.7 | 0.977 6 | 19.3 |
| 40~44 | 8.4 | 4.1 | 0.970 7 | 4.0 |
| 和 | 426.0 | 207.9 | | 204.8 |
| 和×5 | 2 130.0 | 1 039.5 | | 1 023.8 |

资料来源:Martin, Hamilton, Ventura, Menacker, & Park (2002); National Center for Health Statistics (2003).

注:最年轻的和最年长的年龄组,分别包括 15 岁以下和 44 岁以上的生育。分年龄的女婴生育率,是用每出生 1 000 名女婴就有 1 049 名男婴出生的性别比估计的。

## 生殖力量度

经典的生殖力量度都集中在一个性别,一般都在女性人口上。粗生育率(GRR)的计算和解释与总生育率的计算和解释类似,唯一的例外是总生育率加总的分年龄组的生育率只是女性的生育,而非所有的生育,因而它有时被称为“产妇率”而非生育率。如果得不到分年龄组产妇率,那么它们经常用分年龄组的生育率乘以所有出生中女性的比例来估计。在下列公式中右侧的上标是出生的性别。

$$ \begin{aligned} \text{GRR} &= n \sum {}_nF_x^f = n \sum {}_nF_x \times \frac{B^f}{B} \\ &= 5(49.4 + 112.3 + \cdots + 8.4) \times \frac{1\ 000}{2\ 049} \\ &= 1\ 039.5 \end{aligned} $$

净生殖率(NRR)通过将分年龄组的产妇率乘以新生存活到她们的母亲年龄的女儿的比例，$_nL_x^f/_nl_0$，将死亡率也列入考虑之列。该比例是从女性生命表【Life Table】中导出的。具体结果列在了表 1 中，2000 年美国的 NRR 为 1 023.8。

$$\begin{aligned} \text{NRR} &= n\sum {}_nF_x^f \times \frac{{}_nL_x^f}{nl_0} \\ &= 5 \times [(24.1 \times 0.990\,2) + (54.8 \times 0.987\,9) + \cdots] \\ &= 1\,023.8 \end{aligned}$$

因此，2000 年观察的这一分年龄组 1 000 名美国妇女的同期群预期在她们的生命期内可生育 2 130.0 个孩子。其中，1 039.5 个是女儿，并有1 023.8个可活到她们母亲的年龄。

每一个人口都可以用一张分年龄的生育和死亡率的明细表，在某一个时刻加以描述。如果这些比率是无限期继续的，那么，其结果将是一个稳定的以出生率、死亡率、增长率和年龄结构为标记。这些比率都是分年龄的至关重要的比率明细表所“固有的”，因而它们都不取决于原来的人口年龄结构。

## 标准化

因为生育率和死亡率随年龄组的变化而呈很大的变化，因此，如果人口的年龄结构不同，那么这就有可能误导跨人口的粗生育率和死亡率的比较。直接和间接标准化技术使我们得以更为恰当地进行跨人口比率的比较，不过前一个路数较后一个更为可取(Shryock, Siegel, & Associates, 1973)。

粗的人口比率可重写为分年龄组的每一年龄组的比率和人口比例的乘积的和。例如，人口 $i$ 的死亡率便可写作

$$\text{CDR}_i = \frac{{}_\omega D_{0i}}{{}_\omega N_{0i}} = \sum \left(\frac{{}_nD_{xi}}{{}_nN_{xi}}\right)\left(\frac{{}_nN_{xi}}{{}_\omega N_{0i}}\right)$$

人口 $j$ 的直接标准化率($\text{DSDR}_j$)是将人口 $j$ 的分年龄组的比率乘标准或参照人口 $i$ 的分年龄的人口比例求得的：

$$\text{DSDR}_j = \sum \left(\frac{{}_nD_{xj}}{{}_nN_{xj}}\right)\left(\frac{{}_nN_{xi}}{{}_\omega N_{0i}}\right)$$

然后，直接标准化的死亡率就可以和标准化人口的粗死亡率进行比较，因为它们都是以相同的年龄结构计算的。在比较两个人口的比率时，“标准”人口的选择在很大程度上是有些武断的，尽管它一般都优于用更为常规的年龄结构进行的比较。然而，有时惯例会主导标准人口的选择。例如，人口普查局近来宣布，2000 年的美国人口将成为涉及美国人口的历史比较的标准。

任何粗率和直接标准化率之间的差($\text{CDR}_i-\text{DSDR}_j$)都是以标准人口的年龄构成加权的分年龄组率的差的和。我们也可以计算相应的人口 $j$ 的直接标准化率。然后，可将两个人口($\text{CDR}_i-\text{CDR}_j$)的粗死亡率之间的原始差分解成三个部分：归结于率差的部分($\text{CDR}_i-\text{DSDR}_j$)、归结于年龄结构的差($\text{CDR}_i-\text{DSDR}_i$)的部分和一个交互项($\text{DSDR}_j+\text{DSDR}_i-\text{CDR}_i-\text{CDR}_j$)。有些学者则将率和年龄部分之间的交互项分开(Preston, Heuveline, & Guillot, 2001)。

直接标准化和它的推论，两个率(或均值)之间的差的分解是一种在经典人口学有着广泛应用性的技术，其关注的主要问题是

生育率或死亡率。只要一个率是可以被概念化为特定的 $i$ 的率和沿维度 $i$ 的相对比例的积的总和，我们就可以使用它。我们也可以用两个感兴趣的维度来将率标准化（和将率的差分解）（参见比率标准化【Rate Standardization】）。

## 不平等和隔离的量度

有些人口学量度和计数评定和比较跨分析单位的计数分布。例如，收入（以挣得的美元数计数）在一个人口的各个家庭（分析单位）之间的分布是不平等的。这样一些变量的分布常用“份额”来描述。份额的计算要先按计数变量的值将分析单位从最低到最高排列，然后再将分析单位分成等容量的组，一般为 5 组或 10 组。例如，在 2000 年，所有美国家庭中的最穷的五分之一，其收入只占所有美国家庭总收入的 3.6%，而最富的五分之一的家庭则占到了 49.7%。

有时，一个人口（人的计数）可以按一个变量的类别（如地理单位）来描述它的不平等性。非类同指数（$D$）常被用来进行跨一个类别变量的两个或两个以上人口的分布比较。相异指数用下面的公式来计算

$$D = \frac{1}{2}\sum_i \left( \left| \frac{y_i}{\sum_i y_i} - \frac{x_i}{\sum_i x_i} \right| \right)$$

例如，分类变量可能是指美国的四个主要地区（东北部、中西部、南部和西部）；$y$ 是非裔美国人，$x$ 是白人。2000 年，生活在四个地区的每一个的非裔美国人的数目（以 1 000计）分别为 6 100，6 500，18 982 和 3 077；而对应的白人则分别为 41 534，53 834，72 819 和 43 274。相异指数的值为 0，说明跨四个地区的非裔美国人和白种人的分布相同，而相异指数的值为 1，则说明两种人口完全生活在不同的地区。相异指数为 0.20 说明，大约有五分之一的非裔美国人或白种人必定转移到了另一地区，以使两种分布完全一致。尽管相异指数很容易计算和解释，但它却有几个不足之处。例如，它对类别的数目很敏感。其他系数，比如基尼系数【Gini Coefficient】（或基尼比率）或熵指数（$H$）都可作为它的替代计算，不过它们同样也有不足之处。有关这些备择方法的讨论，请参见西格尔的著作（Siegel，2002）。

——Gillian Stevens

（沈崇麟译校）

## 参考文献


Martin, J. A., Hamilton, B. E., Ventura, S. J., Menacker, F., & Park, M. M. (2002, February 12). *Births: Final data for 2000* (National Vital Statistics Report, Vol. 50, No. 5). Hyattsville, MD: National Center for Health Statistics.

National Center for Health Statistics. (2003). *United States life tables*, 2000. Retrieved from www.cdc.gov/nchs/data/lt2000.pdf.

Preston, S. H., Heuveline, P., & Guillot, M. (2001). *Demography: Measuring and modeling population processes.* Oxford, UK: Basil Blackwell.

Shryock, H. S., Siegel, J. S., & Associates. (1973). *The methods and materials of demography.* New York: Academic Press. (Condensed edition by Edward G. Stockwell.)

Siegel, J. S. (2002). *Applied demography.* San Diego, CA: Academic Press.

## 从属访谈(Dependent Interviewing)

从属访谈是在某些纵贯研究【Longtitudinal Research】中所采用的一种方法。例如,定群【Panel】抽样调查中,同样的个人在以后的某些时点还会被重新访谈,这就是一种纵贯研究。根据抽样调查的设计,可以每季度、每年、每两年,或者一段确定的时间之后进行访谈。从属访谈可以被定义为一个过程,在这个过程中,当前这轮调查中问及被调查者的某个问题源自此前某轮调查中某位被调查者的报告(Mathiowetz & McGonagle, 2000)。例如,在某次抽样调查访谈中,一位被调查者的报告称其职业是巴士司机,在后面的访谈中,从属访谈就会采用这位被调查者对其职业的回答。从属访谈不会简单地在两个调查时点都去问被调查者"你现在的职业是什么?"而可能会问,"去年您告诉我们您是巴士司机,现在还是吗?"这位被调查者可以认可此前的回答,或者在情况有变的时候做出不同的回答。

抽样调查数据中常见的误差类型也会出现在从属抽样调查中,包括回忆误差、访员效应,以及在编码过程中的"噪声"。纵贯数据收集的是诸如劳动力市场历史或者收入的持续记录,会遇到一个额外的测量问题,也就是所谓的接缝效应或接缝问题(Doyle, Martin, & Moore, 2000; Lemaitre, 1992)。如果被调查者进行报告的两次抽样调查期间的接缝中出现了不自然的、高水平的可观测行为变化,这时就出现了接缝效应。被调查者倾向于报告他们在以前的访谈中所报告的事情,将其引入当前访谈的时期,这种做法称作前伸(telescoping, forword),它会在两次被调查期间的接缝时间处造成堆积。采用从属访谈设计正是为了减少这些效应,生成长时间上更一致的数据。随着计算机辅助访谈技术的到来,从属访谈在数据收集中更为常见。从技术上而言,使用计算机辅助访谈技术使得把以前报告过的某次访谈的信息前向馈送给下一次访谈是可行的。从属访谈在数据质量上也具有优点,它减少了被调查者和访员的负担,也降低了数据收集和处理的费用。尽管如此,也还是有些问题要考虑。根据不同的实施方式,从属访谈可能会导致低估总体内部真实的变化水平,因为被调查者更易于认可他/她此前的报告。

——Heather Laurie<br>Mike C. Merrett<br>(杨可译　沈崇麟校)

### 参考文献


Doyle, P., Martin, B., & Moore, J. (2000). *Improving income measurement*. The Survey of Income Program Participation (SIPP) Methods Panel. Washington, DC: U. S. Bureau of the Census.

Lemaitre, G. (1992). *Dealing with the seam problem* (Survey of Labour and Income Dynamics Research Papers 92-05). Ottawa: Statistics Canada.

Mathiowetz, N., & McGonagle, K. (2000). An assessment of the current state of dependent interviewing in household surveys. *Journal of Official Statistics*, *16*(4), 401-418.

## 相依观察值(Dependent Observations)

相依观察值是那些以某种方式互相关联或聚合的观察值,它们是那些彼此有一种系统关系的观察值。聚合通常源于研究或抽样设计的策略,聚合可以发生在空间中、时间上或者时空二者。

单次抽样或多阶整群抽样会产生空间上聚合的样本,通常是在地理意义的空间中。识别和选择已知、已定义的聚合并从所选的聚合中挑选一些观察值,会生成彼此不独立的观察值。这些观察值由聚合联系起来,无论它是一个地理上的实体(如城市)、组织(如学校)或是家户。

纵贯研究设计生成时间上连接的观察值。例如,尽管个体是随机选择的,研究者还是可以对每个个体 $i$ 在时点 $t$ 获得一个观察值(如 $attitude_{it}$)。这些合成的观察(如 $attitude_{i1}$ 和 $attitude_{i2}$)互相不独立,因为它们都隶属于 $i$ 这个人。

相依观察的出现违背了一般线性模型的一个关键假设:误差项相互独立且同分布。空间或时间上相连接的观察值通常比独立的观察值更同质。因为那些相连接的观察值不能像独立选择的观察那样提供较多的"新"信息,因而得到的样本特征是异质性较低的。在定性研究中,这意味着研究者在做推论时需要谨慎(King, Keohane, & Verba, 1994)。在定量研究中,这意味着估计值的标准误差会缩小(Kish, 1965)。

如果抽样权数是自变量的函数,因而也是误差项的函数的话,那么我们就应该使用抽样权数,以对整群抽样方法产生的抽样结果作出纠正。此外,温希普和拉德维尔(Winship & Radbill, 1994)建议采用标准误差的白异方差性一致的估计量(white heteroskedastic consistent estimator),而不是试着明白无误地去为误差的方差结构构造建模。这个估计量也可被用来修正样本中的多重性,在标准的统计软件包中我们均能找到这个估计量。

要修正时间上相依的观察,研究者通常要先对时间上自相关的顺序进行评估并建模,进而引进滞后的内生和/或外生变量。在时间序列的文献中,一直都在对这种技术进行广泛而深入的研究。

现在,研究者已经不只是简单地对相依进行纠正,而是越来越多地考虑如何为相依性建模,以及如何对之作出解释。研究者的主要兴趣在于,个体、组织和诸如这样的单位的集群之间的连接。几十年来,网络理论家和分析家一直致力于描述和解释某个或几个聚合中观察到的相依的本质;而近年来的发展更进一步拓展了这些方法,包括了从属网络(Skvoretz & Faust, 1999)和研究时间网络的方法(Snijders, 2001)。最近,布雷克和劳登布什(Bryk & Raudenbush, 1992)发展出了一个给聚类,或者称为"嵌套"数据建模的框架。对这些层级(非)线性模型【Hierarchical (Non) Linear Models】(也被称为混合模型,因其通常含有固定和随机两种效应)而言,在多层次分析【Multilevel Analysis】中的跨层互动,可以对情景效应进行考察:较高层级的单位(学校、组织)会如何影响较低层级的关系。

——Erin Leahey

(杨可译　沈崇麟校)

## 参考文献


Bryk, A., & Raudenbush, S. (1992). *Hierarchical linear models*. Newbury Park, CA: Sage.

King, G., Keohane, R. D., & Verba, S. (1994). *Designing social inquiry: Scientific inference in qualitative research*. Princeton, NJ: Princeton University Press.

Kish, L. (1965). *Survey sampling*. New York: Wiley.

Skvoretz, J., & Faust, K. (1999). Logit models for affiliation networks. *Sociological Methodology*, *29*, 253-280.

Snijders, T. (2001). The statistical evaluation of social network dynamics. *Sociological Methodology*, *31*, 361-395.

Winship, C., & Radbill, L. (1994). Sampling weights and regression analysis. *Sociological Methods & Research*, *23*(2), 230-257.


## 因变量(Dependent Variable)

任何受另一个称为自变量【Independent Variable】的变量因果影响的变量,都可看成是因变量。因变量这一概念在实验研究和非实验研究中有所不同[参见因变量(实验研究中)【Dependent Variable (in Experimental Research)】;因变量(非实验研究中)【Dependent Variable (in Nonexperimental Research)】]。

——Alan Bryman

(沈崇麟译校)

## 因变量(实验研究中)(Dependent Variable [In Experimental Research])

一个因变量是一个量度。在设计精良的实验研究中,因变量是在参与者回应上的自变量【Independent Variable】效应的量度。量度之所以被称为因变量,是因为我们预料它将受到(或取决于)自变量中系统变差的影响。若实验设计得当,因变量中观察到的变差被归结于自变量变差的效应,因而是原因归因之根本。

严格地讲,在非实验研究中,只要操控是不系统的,那么所有的变量都被认为是不独立的。因此,在相关研究中,所有的变量都是相依的。因为这样的研究是无操控的。在多元回归预测研究中,相依量度(dependent measure)被用作预测因子,因而有时也称为自变量(预测因子)变量。在某些研究传统中,如果预测因子是显然不可能作为结果的,那么有时它们也从因果关系上进行解释。例如,一系列环境温度和侵略性的自然观察研究显示,二者具有中强的正关系。但(常识告诉我们)根本不存在人们的侵略性影响了温度这种可能,于是这种相关便被解释为是高温使人们变得有侵略性。这种关系,是以常识或理论为根据的,常常作为因果关系来解释,但这种推论只是一种假设,不如在实验环境中做得那么清楚,因为那些与关键变量共生的外生变量结果有可能被证明才是真正的因果因素。例如,在高温-侵略性的例子中,有可能是在炎热的天气中,人们为了逃避家中的高温外出跑到街上这种倾向才是侵略性真正的唆使因素。(那些不快活的)人群可能是方程中的必要成分而已。

因变量可以取许多形式。因变量的最高自觉形式是自我报告。在这种形式的因变量中,参与者自行表达自己的态度、信念或感情,表明自己对一种现象或他人的行动的因果归因。对这种形式的量度作解释必须加倍小心,其原因至少有二:第一,人们一般都不愿意报告在社会上不受欢迎的行为或态度。一个用来评估令人讨厌的行为或信念的因变量有可能得到社会赞许,但并非真实的回答。另一个更加有害的问题是参与者倾向于以一种帮助研究者证实研究假设的方式来回答问题。揣测实验者的需要这样一种过度合作的回应(参见需求特性【Demand Characteristics】)有可能误导(实验),致使最后证实了一个错误的假设。

与自觉回答连续统相反的是在回答人不知道自己在被观察的情况下观察收集到的那些因变量回答。这种不引人注目或不起反应的量度方法,常在那些正式的实验室环境外进行的研究中使用(参见实验室实验【Laboratory Experiment】)。它的优点是,可为我们提供那些不警觉的参与者真实的或至少是不自觉的回答。不过,与更为直白的自报告量度相比,这样的量度往往是不够灵敏和不太可靠的(Webb, Campbell, Schwartz, Sechrest, & Grove, 1981)。此外,这种量度的道德正当性还存在着很大的争议。正因如此,这样的量度只有在一番深思熟虑之后才可以使用。

在自觉/不自觉的回答维度之外,各种因变量在结构上也有所不同。各种封闭式因变量是高度结构的,它们包含真/假或多种选择答案等方式。这些形式的因变量最适合假设检验研究。在这样的研究中,不仅调查研究的问题是精确设定的,而且需要的信息也是在理论上清楚地设定的。在不是十分正式的产生假设的研究中,这种严格的量度可能不太合适,更为开放的量度可能比较合适。这样的量度涉及研究者的一种可观察的问题的实地记录,无结构的交谈记录等。在这样的情形中,观察的形式是开放的。虽然各种开放的因变量需要用编码和有理论根据的解释来支持研究过程,但它们却可以给我们提供有用的,可发展成为支持专题研究的洞见。

——William D. Crano

(沈崇麟译校)

## 参考文献

Webb, E. J., Campbell, D. T., Schwartz, R. D., Sechrest, L., & Grove, J. (1981). *Nonreactive measures in the social sciences*. Boston: Houghton-Mifflin.

# 因变量(非实验研究中)(Dependent Variable [In Nonexperimental Research])

一般认为因变量在因果上受自变量【Independent Variable】的影响。严格地讲,一个因变量是随机分布的,虽然这种随机性可能会受若干结构因子的影响,因而在最严格的实验【Experiment】研究环境中,自变量是可操控的。然而,在非实验研究中,变量是非操控的,这使得我们有时无法清楚地确定研究的变量中究竟哪一个才是因变量。有时研究者用这一词泛指那些包括在一个隐含因果推论的回归【Regression】方程左边

的任何变量。但是我们必须弄清楚，在不同的研究中究竟应当怎样来确定因变量。在数据来自截面设计【Cross-Sectional Design】时，所有变量上的信息都是共同收集的，因而没有变量是被操控的。这意味着，哪个变量对其他变量有着因果影响这一问题并非立即就可以弄清楚的。尽管我们难以建立因果关系，尤其是在使用截面数据时，但研究者可用几种经验法则来确定把哪一个变量作为因变量。这样的法则有五个，它们未必是互斥的，我们可把它们作为一种指南，在横向和纵向的研究中使用。

### 1.一个变量在时间上领先于另一个变量

甚至在横截面调查中，某个变量记录的事件，其发生的时间早于其他变量，例如，回答人父母的教育获得与回答人的教育获得，后者可被看成是因变量。

在采用纵向研究【Longitudinal Research】方法时，例如，同样本定群【Panel】设计，这个问题会更加复杂一些，因为虽然没有自变量被操控，但数据是在不同的时点收集的这一事实使我们得以使用某种关于时间问题的，因而也是有关因果优先次序的经验手段。严格的实验主义者可能仍然会对此持有异议，他们认为，缺乏操控会对因果优先次序问题造成疑点。为了解决这一难题，同样本追踪数据的分析者，采用了一种称为交叉滞后同样本追踪分析法（cross-lagged panel analysis）来确定在两个或更多个时点量度的成对变量之间的因果优先次序。在这样的分析中，因变量和自变量之间的那条直线，在某种意义上是模糊不清的，因为那个被认为在因果上领先的变量，在第一个回归方程中被作为自变量处理，但在第二个回归方程中则作为因变量处理。相反那个被认为在因果上居后的那个变量，在第一个回归方程中却被作为因变量处理，而在第二个方程中则作为自变量处理。从诸如这样的分析判断来看，“因变量”的确定似乎是相当主观的，无非就是清晰的思路和仔细分析而已。

### 2.来自顺序清楚的变量

例如，某人生命历程【Life Course】的时间，例如，第一次性交和第一次生育就有着清楚的因果次序。

### 3.某一事件在另外一个事件启动前冻结了

一个人给母校的捐赠量取决于他（她）得到的学士学位的类型。在完成学位学业之前捐赠的行为是不会开始的。

### 4.一个变量比另一个变量更容易变化

举一个简单的例子，设想我们在做一个性别和政治保守主义量表关联的调查。人的性别是难以变化的，但是政治取向是可以变化的，似乎我们有理由假设政治保守主义是因变量。当然，可能存在影响性别和保守主义之间的因果路径的中介变量【Interv-ening Variable】或调节变量【Moderating Variable】，但是基本的问题仍然是，对这样一些变量进行因果推论依然是可能的和明智的。

### 5.一个变量比另一个变量更多产

更多产（即更有可能产生成果）的那个变量，譬如说是宗教信仰可以用作自变量；而不那么多产的变量，譬如说某人的孩子数则可作为因变量。

研究者在采用非实验设计（如调查【Survey】设计）时，为了弄清哪些变量是因变量，必定会涉及因果推论问题。实质上，

这一过程就是将来自包括上述五条经验法则在内的，我们对于社会世界的性质的理解和来自现存的理论和研究，以及那些在引起我们注意的领域中的分析方法的运用常识和判断力进行的推论（commonsense inferences）融合。位于那些与交叉滞后同样本追踪分析关联的方法，如因果模型【Causal Modeling】法和路径分析【Path Analysis】背后的正是这种因果推论过程。

在因果次序优先不太明显的时候，因果推论会变得更加困难。例如，如果我们发现在一家大公司的调查结果中，员工日常工作水平和他们的组织公民行为呈负相关，不过因果关系的方向可能是很难确定的，尽管与这些变量关联的研究发现和理论可以给我们提供有益的线索。这种情况可能是从事日常工作的员工较少在公司面前展现自己。不过情况也可能会是那些显示出组织公民行为的员工不太可能被分配从事日常工作。同样本追踪调查常被作为一种解开原因与结果的工具，在这样的情况中使用这一工具，不失为一种明智之举。

——Alan Bryman

Tim Futing Liao

（沈崇麟译校）

## 参考文献

Davis, J. A. (1985). *The logic of causal order* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-055). Beverly Hills, CA: Sage.

Lieberson, S. (1987). *Making it count: The improvement of social research and theory*. Berkeley: University of California Press.

Rosenberg, M. (1968). *The logic of survey analysis*. New York: Basic Books.

# 描述性统计（Descriptive Statistics）

参见单变量分析【Univariate Analysis】。

# 设计效应（Design Effects）

设计效应表明样本设计对估计值方差【Variance】的影响。这种影响的量度涉及从等样本量的简单随机抽样【Simple Random Sampling】求得的等价的估计值方差。因此，一个总体【Population】参数 $y$ 的估计值的设计效应可写作：

$$\mathrm{DEFF}(\hat{y}_{\mathrm{Com}}) = \frac{V(\hat{y}_{\mathrm{Com}})}{V(\hat{y}_{\mathrm{SRS}})} \qquad (1)$$

式中的$\hat{y}_{\mathrm{Com}}$是在讨论中的复杂样本 $y$ 的估计值，$\hat{y}_{\mathrm{SRS}}$是等样本量的简单随机样本的估计值。

设计效应针对估计值和样本设计二者。因此，来自同一调查不同数量的估计值可能会有不同的设计效应，好比来自不同调查【Survey】同一数量的估计值。设计效应上的重要影响有三个：样本集群（sample clustering）、样本分层（sample stratification）

和样本包含概率(sample inclusion probability)。

多阶整群样本设计(参见多阶抽样【Multistage Sampling】)一般会导致设计效应大于1。这是因为集群单位内样本单位,总地讲都比总体单位相互之间更为相似(也就是集群是比较同质的)。情况也许的确如此,例如,居住在小区域内的住户,户中的学生往往都在同一所学校上学,户中的人们也常常都在同一个雇主处从业等。因集群而引起的设计效应,取决于集群内的同质程度和(平均的)集群样本的大小:

$$\mathrm{DEFF}(\hat{y})_{\mathrm{Clus}} = 1 + (b-1)\rho \qquad (2)$$

式中的 $b$ 是(平均的)集群样本的大小,而 $\rho$ 是群间相关【Correlation】系数。

比例分层(参见分层样本【Stratified Sample】)的设计效应一般都小于1。无独有偶,这是因为层都是比较同质的。但是与集群不同,分层免于样本或为集群的一个子集,确保样本能代表所有的层。分层和目标变量的相关程度越高,因分层而引起的设计效应就越小(越有利)。

因变量包含概率(包括选择概率和回答概率)而引起的设计效应一般都大于1,且它越大,概率中的变差也越大。在层的总体的方差已知或可估计时,为了提高估计的精度而将最优配置施予层时,会出现显著的例外。因变量包括概率引起的设计效应可写作:

$$\mathrm{DEFF}(\hat{y})_{\mathrm{Inc}} = \frac{n}{N^2 S^2}\sum_{h=1}^{H}\frac{N_h^2 S_h^2}{n_h} \qquad (3)$$

式中的 $n_h$,$N_h$ 和 $S_h^2$ 分别是在总体被分为 $H$ 个互斥而穷尽的类时,$h$ 类样本的大小、总体的大小和总体方差,而 $n$,$N$ 和 $S^2$ 则分别是整个样本的大小、总体大小和总体方差。

许多技术报告都刊印了一系列估计值的设计效应。这些信息对于今后的调查设计,以及来自当前调查结果的解释都很有用处。

一个与之密切相关的概念是设计因子,它表明样本设计对估计值的标准误差的影响,因此也表明了对置信区间【Confidence Interval】宽度的影响:

$$\mathrm{DEFF}(\hat{y}_{\mathrm{Com}}) = \frac{s.e.(\hat{y}_{\mathrm{Com}})}{s.e.(\hat{y}_{\mathrm{SRS}})} = \sqrt{\mathrm{DEFF}(\hat{y}_{\mathrm{Com}})} \qquad (4)$$

——Peter Lynn

(沈崇麟译校)

## 参考文献

Kish, L. (1965). *Survey sampling*. New York: Wiley.

Kish, L. (1995). Methods for design effects. *Journal of Official Statistics*, *11*(1), 55-77.

# 行列式(Determinant)

一个方阵的行列式是与那个矩阵有关联的唯一的标量。矩阵 **A** 的行列式常常被表示为行列式 **A** 或 $|\mathbf{A}|$。行列式符号以垂线代替矩阵的括号。因此,对于

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, \quad |\mathbf{A}| = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}$$

行文中有时将

$$|\mathbf{A}| = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}$$

称为行列式函数,而将产生的标量作为行列式的值。1 阶和 2 阶矩阵的行列式计算是简单明了的。如果 $\mathbf{A} = [a_{11}]$,那么 $|\mathbf{A}| = |a_{11}| = a_{11}$。如果

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

那么

$$|\mathbf{A}| = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21}$$

注意,上述二阶行列式的例子是主对角线上的元素的乘积减去对角线外的元素的乘积。一种称为行(或列)扩张的方法可以用来系统地计算 3 阶或 3 阶以上的矩阵的行列式。扼要地讲,对 $n$ 阶矩阵 $\mathbf{A}$ 而言,行扩张将产生任何一行 $i = 1, 2, \cdots, n$ 的等式,

$$|\mathbf{A}| = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} |\mathbf{A}_{ij}|$$

列扩张将产生任何一列 $i = 1, 2, \cdots, n$ 的等式,

$$|\mathbf{A}| = \sum_{i=1}^{n} (-1)^{i+j} a_{ij} |\mathbf{A}_{ij}|$$

对于任何一列 $j = 1, 2, \cdots, n$。在这两个等式中,$|\mathbf{A}_{ij}|$ 是由减去 $\mathbf{A}$ 的第 $i$ 行和第 $j$ 列形成的子矩阵的行列式,常被称为 $\mathbf{A}$ 的小 $a_{ij}$。因此,3 阶矩阵 $\mathbf{A}$ 的第一行的扩张是

$$|\mathbf{A}| = \sum_{j=1}^{3} (-1)^{i+j} a_{ij} |\mathbf{A}_{ij}|$$
$$= a_{11} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12} \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} +$$
$$a_{13} \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}$$

有几条可用于行列式的行与列处理的定理,它们对行列式的计算很有帮助。这几条定理是:

- 矩阵 $\mathbf{A}$ 的行列式等于 $\mathbf{A}$ 的转置矩阵的行列式(即 $|\mathbf{A}| = |\mathbf{A}^{\mathrm{T}}|$)。
- 交换一个行列式的任何两行都要改变它的符号。
- 如果一个行的每一个元素都乘以因子 $k$,那么行列式的值也要乘以 $k$。
- 如果将一行的数量的倍数加到另一行,那么行列式的值不变。

如果矩阵的行列式是非零的,那么该矩阵便是非奇异的,因而将会有一个逆。行列式将被用来计算矩阵的逆,然后二者都将被用于求解方程系统,诸如这样的系统可见于回归【Regression】分析中。行列式也常和 Cramer 法则一起作为一种求解联立【Simultaneous】线性方程【Equation】,以及确定函数【Function】的极端点是否最大或最小值的方法使用。

如果矩阵的行列式等于零,那么我们就说矩阵是奇异的。这表明两个或更多的行或列线性相依【Linearly Dependency】。在代数中,这意味着矩阵的逆是不能计算的。从实际情况看,这就意味着用矩阵表示的模型中的两个或更多个变量是彼此的线性变换【Transformations】。

——Timothy M. Hagle

(沈崇麟译校)

## 参考文献

Cullen, C. G. (1990). *Matrices and linear transfor-*

*mations* (2nd ed.). New York: Dover.

Pettofrezzo, A. J. (1978). *Matrices and transformations*. New York: Dover.

## 决定系数(Determination, Coefficient of)

参见 *R* 方【*R* Squared】。

## 决定论(Determinism)

决定论通常是带着轻蔑来使用的术语,用以描述某个论点或方法把因果关系过度简化为一组因素活动所引起的结果。因此,举例来说,经济决定论把社会、政治和文化效应都归因为与经济的关系,认为这才是根本的原因;技术决定论也赋予了技术革新以同样的地位;文本决定论将经过媒介传递的信息对个人和受众的影响归因为信息或媒体的文本性质;诸如此类。与此相关但不同的一个术语是还原论,人们也批评它把一套复杂的环境简化为了一组有限的特征,且假定该组特征显要。

经常被当作决定论来批评的方法论是一些主张采用整体化的方式来解释的方法论,例如,马克思主义、弗洛伊德主义、信息主义、文本主义、系统论等;还有一些想要把诸如天气、阶级、性别、种族或地理环境等自然、社会或物理特性视为社会活动之根本的方法论;以及那些从人类活动中拣出诸如技术、性别、工作等某些特别的方面,认为它们对人类活动其他方面有巨大影响的方法论。

当然,说明决定论者和/或还原论缺陷的学术研究还不够。但在有关决定论的辩论中,常常隐含着研究者在寻求对人类现象进行充分因果解释时所找到的基本议题,即到底这番搜寻是否值得,到底时空中人类的活动领域是否可以被视为整体(或者一套互相重叠的整体),到底在社会因果关系中是否存在系统的层级。简言之,有关决定论的争论可能代表了对因果关系的争论。

思想家们采用过许多不同的方式来探求因果关系。社会计量模型有一个有趣的成果,即对作为因果关系推论的可计量的可预测性所产生的疑问:这番辩论在决定论的数学模型和随机的统计方法之间展开(后者结合了机遇结果)。在元理论的层面,有人主张整体的结构决定其性质及其各部分的功能;系统论就是这种方法的一个典型代表。其他人主张整体是各部分的综合,且其各部分在某种意义上是可知的;实用主义的方法就经常采取这个立场。

第三种方法关注的既不是各部分,也不是整体,而是总是处于冲突中的各种力量,它认为正是这些力量在任意一个具体的时空节点上构造了相对于任意整体的具体各部分。这种冲突在具体的时空中产生结果,决定了各部分之间关系的性质,由此建构了整体。这种方法在20世纪后半叶在不同的学科和不同的整体观下发展起来(例如,教育心理学家让·皮亚杰称其为“运演结构主义”,而对马克思主义的地理学家戴维·哈维来说,就是“辩证的、历史地理学的唯物主

义理论”）。

这种方法把因果关系视为一个不断持续的复杂过程，在这一过程中各种抽象力量互相冲突，在各个具体的历史地理背景中以具体的实质方式来呈现，其结果自身又进入了以后的因果关系的过程之中。它拒绝那些认为各个因素之间具有直接、线性因果关系的观点，但的确试图在决定结果的复杂关系中分辨出能量或支配的层级结构。

哈维（Harvey, 1973/1989）的作品围绕着这些有关城市化的方法论问题，在因果关系的空间和时间维度上进行了一番彻底的探究。《社会公正与城市》（*Social Justice and the City*, Harvey, 1973）是对这个方法论议题的早期的详细探讨；而《后现代的状况》（*The Condition of Postmodernity*）则是他最著名的作品。

——Chris Nash

（杨可译　沈崇麟校）

## 参考文献


Harvey, D. (1973). *Social justice and the city*. Baltimore: Johns Hopkins University Press.

Harvey, D. (1989). *The condition of postmodernity*. Oxford, UK: Basil Blackwell.


# 确定性模型（Deterministic Model）

确定性模型是不包含随机成分的模型。

18 世纪初，确定性模型开始广泛用于物理过程的研究，它的应用得益于数学家雅各布·贝尔努利、约翰·贝尔努利和莱昂哈德·欧拉等人开发的微分方程。把给出的一个变量值作为它的一个极小的早期时点值的函数的微分方程，仍然是确定性模型【Model】最常见的形式。已经证实，微分方程能用于各种各样的大气和机械现象，并已产生了大量数学文献。有限差分方程变量的值是它的滞后某个有限值（一般为 1）时期的值的函数——是确定性模型的另一种形式。不过它在数学上的优越性不如微分方程。差分方程被普遍用来为变化较慢的生物和社会过程建模。这样的过程有生物总体的跨代际过程或经济过程中跨预算年的过程等。

倘若函数的形式、模型的系数值和过程的初始值已知，那么确定模型的值就完全是可预测的。在 18 和 19 世纪，物理世界就是由许多自然科学根据完全确定的法则做出的假定运转的。这就意味着，至少在理论上意味着物理世界是完全可以预测的。法国数学家皮埃尔·拉普拉斯在他 1812 年发表的著作《分析概率论》（*Analytic Theory of Probability*）中一段被广泛引用的简要陈述中说道：

“立即给人以一种智力，它可以理解所有自然赖以有生命的所有力量和组成自然的生命的各个方面……它将在同一公式中包含宇宙中最大的物体和最小的原子的运动。对它而言，没有什么是不确定的，未来犹如过去一样呈现在它的眼前。”（Mackay, 1977: 92）

诚如拉普拉斯所言，确定性模型的一个奇妙的特性是，它不仅可以从模型预测未来，还可以推断过去的行为。例如，天文学

家因此而用决定性方程为新发现的彗星建模,不仅能发现彗星将去向何方,也能发现它来自何处。同样,一枚弹道导弹路径的确定性模型可用于决定它的近似的起飞点,即使起飞是未曾被观察到的。

在 20 世纪,确定性宇宙这一概念遇到了三个挑战。第一个挑战发生在 20 世纪初,由于量子力学理论的发展,人们发现在原子水平,物理过程有着本质上的随机成分,它与早先设想的完全确定性的世界相左。第二个挑战发生在整个 20 世纪,概率论和统计学(包括随机的【Stochastic】过程研究)的数学领域把随机性也包括在内,发展到了一个可与 19 世纪确定性过程媲美的高深的水平。随后在社会科学中进行的工作,理所当然地把注意力都主要放在那些清晰地设定了随机成分的模型上。

第三个挑战是在 20 世纪末,混沌过程开始受到广泛的关注(参见 Kiel & Elliott, 1996)。混沌模型是确定性的——它们一般都是非线性的差分方程,但它产生的值的序列都是随机的,且这些值在一个长时期是不可预测的,除非初始值是可以无限精确地被估计的,而实际上这是不可能的。

尽管受到了这些挑战,确定性模型在许多场合还是很有用的。尽管面临量子力学和混沌理论的挑战,但确定性方程还是能为那些发生在原子水平之上的许多物理过程提供极为精确的预测。在近似社会过程方面,确定性模型也是很有用的,即使在那些已知含有随机组成成分的场合,情况也同样如此。这方面的例子包括人口、预算和国际军备竞赛的增长。

——Philip A. Schrodt

(沈崇麟译校)

### 参考文献

Kiel, L. D., & Elliott, E. (1996). *Chaos theory in the social sciences*. Ann Arbor: University of Michigan Press.

Mackay, A. L. (1977). *Scientific quotations*. New York: Crane, Russak.

## 去趋势(Detrending)

时间序列【Time-Series】的数据常常会展现一种随时间上升或下降的趋势。一种趋势可能表现为一种平均值和/或方差【Variance】随时间变化的方式。一张时间序列的曲线图可以是线性的,展示随时间变化的趋势均值。将这样的变量在另一个变量上做回归,常常会产生一个较高的 $R$ 方【$R$-Squared】,但估计的参数值却并没有反映出自变量和因变量之间的关系。不仅如此,估计的参数值将会向上偏倚【Bias】,因为这种趋势会导致两个变量之间的伪关系【Spurious Relationship】,反映存在于它们二者之中的共同趋势。如果情况确实如此,那么为了得到自变量和因变量之间的真实关系,在拟合模型之前观察必须先去趋势。图 1 显示了一个在时间上存在某种趋势的系列及在去除趋势之后的同一系列。

解决这一问题的最简单的办法是在回归【Regression】分析中,引进一个时间或趋势变量作为自变量【Independent Variable】。趋势变量不必是线性的,如果系列展示一种

立方的趋势，那么趋势变量可以是 $X^3$，在这里 $X$ 是时间。此外，有些时间系列展示周期性的趋势，这时比较恰当的做法是也引进一个季节性趋势变量。在方程的右边引进一个趋势变量，将会产生它与因变量之间的真关系的系数估计值。

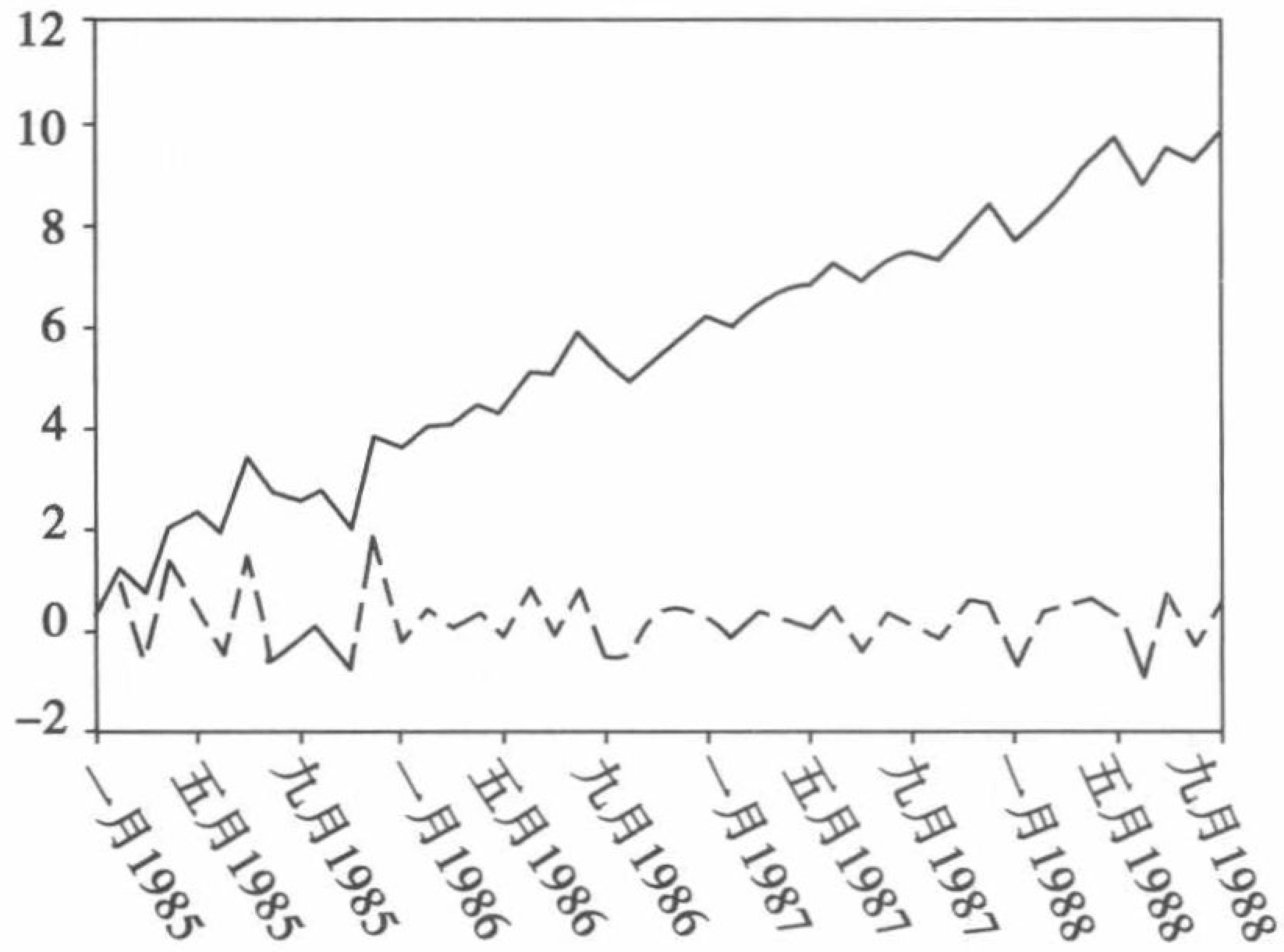



**图 1　趋势系列和去趋势之后的同一系列例解**

如果因变量和自变量展现出不同的趋势，比如说，因变量呈一种线性的趋势，而自变量是二次的，那么采用一种更为复杂的去趋势过程可能比较有用。首先，将因变量在时间上做回归，得到残差【Residuals】$\hat{u}_{1t}$。然后再做自变量对时间的回归，得到残差 $\hat{u}_{2t}$。最后做 $\hat{u}_{1t}$ 对 $\hat{u}_{2t}$ 的回归。由于时间趋势已从两个变量上取出，$\hat{u}_{2t}$ 上的系数将是因变量和自变量之间的真关系。如果我们想要报告的不仅是因变量和自变量之间的关系，而且也想报告它们独特的时间趋势，那么这种去趋势的方法就会格外有用。

以这种方式去趋势只适用于研究那些，其趋势是确定性的系列。如果趋势是随时间变化的，比如说是随机的【Stochastic】，这时 ARIMA（自回归移动平均）建模法便是一种恰当的用来发现潜在关系的方法。单位根【Unit Root】检验可能是用来确定系列是否是平稳的比较恰当的方法。在 ARIMA 建模法中，通过差分系列或从时间 $t$ 的变量值减去前值来去趋势。诸如简单线性趋势这样的趋势，通过在回归中引进趋势变量就很容易去除，也可以在 ARIMA 模型中用一阶差分方程或一个因变量 $y$ 是从时间 $t$ 到前一个时间 $t-1$ 的变化的方程来去趋势。

——Andrew J. Civettini

（沈崇麟译校）

## 参考文献

Gujarati, D. N. (1995). *Basic econometrics* (3rd ed.). St. Louis, MO: McGraw-Hill.

Harvey, A. C. (1990). *The econometric analysis of time series* (2nd ed.). New York: Philip Allan.

# 偏差个案分析（Deviant Case Analysis）

偏差个案分析是谈话分析【Conversation Analysis, CA】中的一个步骤，用来检验谈话互动中所提出的现象的有效性和一般性。作为一般性的方法论原则，它可与假设检验和证伪相比。不过，它的某些特点使它成了 CA 中的一个基本要素。

CA 的主要目标就是要发现稳健的谈话序列分析规则，它不仅是统计上观察到的规律，而且也是谈话的参与者作为标准预期明确使用和适应的规律。通过对谈话数据语料库的归纳分析，也许我们可以提出谈话顺序的潜在规则，例如，互致问候，或者接受或拒绝邀请的方式。偏离个案就是那些似乎与这些提出的规则或模式不符的个案。

这些个案的重要性，除了它们不确定的状态之外，还在于它们常常会成为证明参与

者的规范取向的例子,在一定程度上回应了具有早期常人方法论【Ethnomethodology】特征的“破坏性实验”(breaching experiments)。和那些实验不同的是,CA 的偏离个案是在有案可查的谈话的语料库中自然发生的。史蒂芬·克莱曼和道格拉斯·梅纳德(Clayman & Maynard, 1995)对 CA 中偏离个案的用途作了如下三种定义:(a)经过分析,偏离个案可能会与标准个案一样,展现同样的惯常取向;(b)也许据此可以提出一种囊括(明显)偏离个案和主(个案)集的新的分析;(c)偏离个案也可能会展现出其自身的独特顺序,以至更适合将其归入另一个集合。

第二类,也即提出修订分析的一个经典例子,是谢格洛夫(Schegloff, 1968)在早期的 CA 研究中所做的有关电话谈话通常的开始方式的研究。谢格洛夫分析了 500 个例子的语料库,提出了“首先开口的分布规则”,其大意是,诚如所料,那个先开口的人是接电话的。但有一个个案偏离了这个规则。这个个案引发了对该规则的质疑和彻底修改。我们提出的不是一条由接电话的人开启第一轮对话的规则,这样的对话可以被认为是互致问候(如问道“你好”,答道“你好”)的第一个部分;而我们提出的则是一条可以解释所有 500 个个案的修正的规则,这就是所谓的召唤—回应顺序(the summons-answer sequence),在这个规则中,第一声“你好”(或别的什么话)被看成是另一类(召唤—回应)的第二部分,是对电话铃声这样的召唤的回应。

——Derek Edwards

(杨可译 沈崇麟校)

## 参考文献

Clayman, S.E., & Maynard, D. (1995). Ethnomethodology and conversation analysis. In P. ten Have & G. Psathas (Eds.), *Situated order: Studies in the social organization of talk and embodied activities* (pp. 1-30). Washington, DC: University Press of America.

Schegloff, E. A. (1968). Sequencing in conversational openings. *American Anthropologist*, *70*, 1075-1095.

# 偏差(Deviation)

偏差是一个普遍使用的有关某种形式的差异(discrepancy)的术语。一个最简单的例子是一个观察值与总体【Population】均值 $\mu$,即所有可量度的个体的平均数之间的偏差或差别。一个观察值和 $\mu$ 之间的平均(期望)平方偏差就是总体方差【Variance】。一个与之关联的量度是我们所做的观察和样本均值之间的平方偏差,它用样本【Sample】方差来表示。这就是说偏差是以样本均值,而非总体均值定义的。

在更广义的语境中,偏差指的是观察的样本和某个可能代表所有数据产生的过程的模型之间的差异。也许最简单也众所周知的例子是确定观察的样本是否是从某个特定分布中抽出来的,一般我们感兴趣的分布是正态分布【Normal Distribution】。最为普遍的用于量度观察值的分布对假设分布的偏差的量度称为 Kolmogorov 差距(例子请参见

Conover，1980）。这个方法被扩展到量度两个分布之间的总偏差（Doksum & Sievers，1976），它可给我们提供一种令人感兴趣的有关如何进行两个独立的组之间比较的思路（更为详细的介绍和合适的软件，请参见 Wilcox，2003）。

偏差这一概念也出现在其他几种情况中。例如，在回归中，常见的一种模型是 $Y=\beta_0+\beta_1X+\varepsilon$，式中的 $\varepsilon$ 一般被看成是一个有零均值的变量。这样，如果我们得知 $X=10$，那么 $Y$ 的均值等于 $\beta_0+\beta_1X=\beta_0+\beta_1\times10$，一般，式中的 $\beta_0$ 和 $\beta_1$ 都是根据我们可资利用的观察样本估计的未知的参数。实质上，模型假设 $Y$ 的均值和 $X$ 之间存在着线性关联。在这里偏差可能涉及假设的模型和在一个特定的研究中收集的数据之间存在的任何差异。因此，如果 $Y$ 代表儿童的认知功能，而 $X$ 则代表在家中的攻击性水平，那么一个常用的策略是假设这两个量度之间有着线性关联。一个基本的问题是测量这种假设的线性关联和试图评价用来描述数据特性的模型精确性中的真实关联之间的差异。也就是，我们是否有理由假设，对于 $\beta_0$ 和 $\beta_1$ 的某种选择，$Y$ 的均值是否真的等于 $\beta_0+\beta_1X$？测量数据与这种模型偏离了多少的量度已经被设计出来，Kolmogorov 差距就是其中一种。这些量度可以用来进行有关线性模型假定的经验检查（Wilcox，2003）。

另一种情况是在定类数据中出现的偏差。我们来考虑一个最为简单的例子，在这个例子中，我们要求一对成年的夫妇，在一个 5 级量表上给某位政治领袖打分。一个可能的问题是，女性的回应是否与她的配偶的回应是各自独立的？现在，我们的策略就是测量对在独立性（假设）下期望的回应模式的偏离。更为广义地讲，偏差的作用是在考察某个拟建模型的合理性（plausibility）时，对拟合优度作出评估（Goodman，1978）。

——Rand R. Wilcox

（沈崇麟译校）

## 参考文献


Conover，W. J.（1980）. *Practical nonparametric statistics*. New York：Wiley.

Doksum，K. A.，& Sievers，G. L.（1976）. Plotting with confidence：Graphical comparisons of two populations. *Biometrika*，*63*，421-434.

Goodman，L. A.（1978）. *Analyzing qualitative/categorical data*. Cambridge，MA：Abt.

Wilcox，R. R.（2003）. *Applying contemporary statistical techniques*. San Diego，CA：Academic Press.


# 历时的（Diachronic）

历时的特征描述了一种随时间变化的现象，隐含研究者持有时序的、历史的或纵向的观点（参见共时的【Synchronic】）。

——Tim Futing Liao

（沈崇麟译校）

# 日记(Diary)

日记是一种自然的数据收集法,用以获取个人的定期(通常是当天)数据(Corti,1993)。日记更适合用于得到脉络化的、有细节和顺序的信息,且与问卷【Questionnaire】相比,来自日记的数据通常也更加可靠,因为它们把回顾回忆偏差减到了最小,而且它们(是自我管理的)还可以确保匿名【Anonymity】,并防止访员效应【Interviewer Effect】的产生。不过,有可能会有与选择或志愿者偏倚有关联的问题。因此,为了鼓励日记作者在整个期间不间断地写日记,常常会给他们一些奖励。日记也可以和后续的访谈一起使用,以便在访谈时适时插入话题(Zimmerman & Wieder,1977)。

## 日记的形式

日记法有一系列结构形式。

- 非结构式日记没有或几乎没有特定的格式。它们常常都是主动记录的,日记作者自己决定哪些事值得报告。格式随意的日记是自传式的,采用自然语言(或者像 Pepy 的日记那样用代码),更多的是一种定性的资源,而非一种系统的数据收集方式。

- 半结构式日记有研究者设定的特定格式或架构,它们通常都是(研究者)约写的。这种格式的日记不仅可以直接比较,而且日记作者也可以用自己的话来描述事件,并对其加以扩展。得到的数据通常都结合了定性和定量两种模式。

- 结构式的日记更接近于生活史,这是在时间单位固定这一意义上而言的。此外,它们还有两个关键的组成部分:(a)特定的行动是什么(相关的活动通常都是预先设定和预先编码的);以及(b)这个已知的事件或活动持续了多长时间。这可以直接生成可比较的定量数据。最常见的例子就是时间利用日记(Juster & Stafford,1985)。

日记包括的时间长短差异很大。时间利用研究大多只持续几天,而其他形式安排的时间,或要包括重大事件,或要为行为提供稳定的估计值,因而大多需要一星期到一个月的时间。日记法近年来被计算机网络彻底改造了。在线日记变得越来越普遍,并涵盖了上述一系列的日记形式。

## 日记法的应用

家庭调查常常要用到日记,而市场研究者和经济学家则用日记来估计消费者的消费模式。日记在健康和疾病研究(Verbrugge,1980)、饮食与营养研究、酒精与消遣性药物使用研究以及服药依从性研究中也有广泛的应用。至少从 1930 年代经典的金赛研究开始,私人日记就被用于性行为研究了,而近年来,它则被用于男性同性性行为的男性风险行为研究(Coxon,1996)。

——Anthony P. M. Coxon

(杨可译　沈崇麟校)

## 参考文献

Corti, L. (1993). Using diaries in social research (*Social Research Update*, Iss. 2). Guildford, UK: University of Surrey, Department of Sociology.

Coxon, A. P. M. (1996). *Between the sheets: Sexual diaries and gay men's sex in the era of AIDS.* London: Cassell.

Juster, F. T., & Stafford, F. P. (1985). *Time, goods and wellbeing.* Ann Arbor, MI: Institute for Social Research.

Verbrugge, L. (1980). Health diaries. *Medical Care*, *18*, 73-95.

Zimmerman, D. H., & Wieder, D. (1977). The diary-interview method. *Urban Life*, *5*, 479-498.

## 二分变量(Dichotomous Variables)

二分变量是一种在被观察或测量时,取值为仅有的两个可能值中的一个变量。很多时候其变量值表示的是一个测定变量(如年龄:65 岁以下/65 岁及 65 岁以上)或者一个属性(如性别:男/女)。

如果二分变量代表的是另一个变量,那么相关变量既可能被观察到也可能观察不到。例如,一个二分变量可以被用来表示某条法律是否会被通过。这时这个二分变量(通过/未通过)就代表着对这次立法真实的、可观察的投票表决。相反,每个立法委员所投的票则是一种独特和观察不到的在投票赞成或反对这项立法之间的概率分布的结果。

二分变量通常都用 1 和 0 作为两个可能的值来测量。1 和 0 的使用通常都没有与变量相关的特殊含义。我们也可以简单地取 2 和 0,或者 1 和-1 来作为这两个值。尽管如此,1,0 编码在求个案的总和或求平均数时显出了它的优势,因为这就是分值为 1 的个案的比例。1 和 0 如何选取和分配通常都以易于解释和研究问题具体的性质为依据。

二分变量既可以作为因(内生)变量【Dependent (Endogenous) Variable】使用,也可以作为自(外生)变量【Independent (Exogenous) Variable】使用。许多社会科学的研究问题都使用二分因变量。这样问题常常涉及个人的二分选择或结果(例如,买不买,投赞成票还是反对票,确认还是否认)。有二分因变量的估计模型通常需要使用非线性技术,因为线性模型(如普通最小二乘【Ordinary Least Square】)也许在理论层面是不切实际的,而且会生成低效或不一致的结果。

二分变量当作自变量来使用时,常常被称为“控制变量”或“哑”变量【“Dummy” Variable】。遗憾的是,以这种方式这样来称呼二分自变量,可能会导致无法精确合理地解释变量的测量或者它被纳入模型的原因。与任何其他变量一样,二分自变量也应该有坚实的理论基础,并能被恰当地具体化。

值得注意的是,我们不能太急于将变量“二分”(即将可测量的基本变量变成二分的形式),原因有三:第一,这种转换可能在理论上站不住脚。诸如概率单位分析【Probit Analysis】和概率对数【Logit】这样的估计方法的普及,让某些研究者画蛇添足地把连续的因变量转换成了二分的形式。第二,即使理论是合理的,这样的转换也必定会导致信息的丢失。比方说,知道最高法院通过了一项决议是一回事,但知道它是以 9-0通过的,还是以 5-4 通过的则是另外一回事。第三,如果一个变量有重要的剩余类别,把一个变量强行分为两类可能就不那么简单或者合适。例如,如果有人要对使用左右手的习惯进行二分测量,那么双手都灵巧的个体,或者因上肢残障不得不改用另一只手的个体的数量可能会给编码和分析带来问题。

——Timothy M. Hagle

(杨可译　沈崇麟校)

### 参考文献

Hardy, M. A. (1993). *Regression with dummy*

*variables*. Newbury Park, CA: Sage.
Maddala, G. S. (1983). *Limited-dependent and qualitative variables in econometrics*. Cambridge, UK: Cambridge University Press.

## 均值差(Difference of Means)

社会科学家常常有兴趣发现两个社会群体是否相同,例如两个社会阶层之间或男女之间的收入水平是否相同。这需要进行显著性检验【Significance Testing】,虚无假设是两个均值$\mu_1$和$\mu_2$之间没有差异。在涉及的均值只有两个时,最常用的方法是一个简单$t$检验【$t$-Test】。这个方法也常用于检验比例差【Difference of Proportions】,而在涉及更多个组时,为了检验组间差异【Between-Group Differences】,我们自然会选择方差分析【Analysis of Variance】。一个与之密切关联的方法是多重分类分析【Multiple Classification Analysis】。还有各种可资利用的处理不同情况的多元比较法,例如,用于关联或相依显著性假检验的 Bonferroni 法【Bonferroni Technique】。总之,均值差检验只是许多进行统计比较【Statistical Comparison】的方法中的一种。

——Tim Futing Liao

(沈崇麟译校)

## 比例差(Difference of Proportions)

在将人们分组时,每个组的不同比例将显示某种性质、态度或者结果。例如,投票给民主党的男和女的不同比例、患乳腺癌的美国人和日本人的不同比例,以及赞同未婚先孕的主流和原教旨新教徒之间的不同比例等。

主流和原教旨新教徒对流产的态度的差别,可以用1990年世界价值观调查的美国部分的数据来估计,它可总结概括为下表:

| | 主流 | 原教旨 | 合计 |
|---|---|---|---|
| 赞成 | 172($n_{11}$) | 23($n_{12}$) | 195($n_{1.}$) |
| 不赞成 | 404($n_{21}$) | 125($n_{22}$) | 529($n_{2.}$) |
| 合计 | 576($n_{.1}$) | 148($n_{.2}$) | 724($n_{..}$) |

注:为了简便起见,我们将"世界价值观"调查作为简单随机抽样【Simple Random Sampling】处理,而实际的调查设计则比较复杂。

在$n_{.1}=576$个主流新教徒中,有$n_{11}=172$个赞成未婚女性流产,比例为$p_1=172/576=0.298\ 6$。在$n_{.2}=148$个原教旨新教徒中,有$n_{12}=23$个赞成未婚女性流产,比例为$p_2=23/148=0.155\ 4$。这样,在被调查的新教徒中,比例差等于$p_1-p_2=0.143\ 2$。调查的主流新教徒可能赞成未婚女性流产的比例比原教旨新教徒高14.32%。

如果我们希望从样本推广到美国新教徒的总体,我们需要知道比例差的抽样分布【Sampling Distribution】。在大样本中,比例差$p_1-p_2$近似正态分布,它的标准差【Standard Deviation】被认为是比例差的标准误【Standard Error】。

不同的标准误被用于置信区间【Confidence Intervals】和假设检验【Hypothesis Tests】。就置信区间而言,最为普遍的标准

误公式是

$$s_{p_1-p_2}=\sqrt{\frac{p_1(1-p_1)}{n_{.1}}+\frac{p_2(1-p_2)}{n_{.2}}}$$

式中的

$$p_1=\frac{n_{11}}{n_{.1}},p_2=\frac{n_{12}}{n_{.2}},$$

相应的$(1-\alpha)\times100\%$的置信区间是

$$p_1-p_2\pm z_{1-\alpha/2}s_{p_1-p_2}$$

式中的 $z_{1-\alpha/2}$是$(1-\alpha/2)$的标准正态分布【Normal Distribution】的第 100 个百分位【Percentile】。

尽管使用非常普遍，但是这种置信区间的涵盖范围不是很理想，特别是在表的计数$n_{ij}$比较低时。更好的涵盖，特别是对低计数，是通过将 1 加到每个 $n_{ij}$和将 2 加到每个$n_{.j}$得到的(Agresti & Cato, 2000)。经过改进的标准误是

$$\tilde{s}_{p_1-p_2}=\sqrt{\frac{\tilde{p}_1(1-\tilde{p}_1)}{n_{.1}+2}+\frac{\tilde{p}_2(1-\tilde{p}_2)}{n_{.2}+2}}$$

该式中

$$\tilde{p}_1=\frac{n_{11}+1}{n_{.1}+2},\tilde{p}_2=\frac{n_{12}+1}{n_{.2}+2}$$

其中，改进过的$(1-\alpha)\times100\%$的置信区间是

$$\tilde{p}_1-\tilde{p}_2\pm z_{1-\alpha/2}\tilde{s}_{p_1-p_2}$$

就新教徒的数据而言，$\tilde{p}_1-\tilde{p}_2=0.139\ 3$，$\tilde{s}_{p_1-p_2}=0.035\ 5$，$z_{.975}=1.96$，于是一个 95%的置信区间，其范围为 0.069 8~0.208 8。换句话说，主流新教徒总体中较可能赞成未婚女性流产的百分比可能为 6.98%~20.88%(有时用这一公式计算的置信限超过$-p_1$ 的下界或 $1-p_1$ 的上界。在遇到这种情况时，都按惯例向内调整置信区间)。

如果我们想要检验等比例的虚无假设【Null Hypothesis】，我们使用一个略有差异的标准误。在这一虚无假设下，我们把调查的新教徒看成是一个 $n_{..}=724$ 个人的单独的样本，其中 $n_{1.}=195$ 个赞成未婚女性流产，比例为 $p=195/724=0.269\ 3$。在这样的设置中，合适的标准误公式是

$$\hat{s}_{p_1-p_2}=\sqrt{p(1-p)\left(\frac{1}{n_{.1}}+\frac{1}{n_{.2}}\right)}$$

式中的

$$p=\frac{n_{1.}}{n_{..}}$$

对新教徒数据而言，$\hat{s}_{p_1-p_2}=0.040\ 9$。合适的检验统计量是

$$z=\frac{p_1-p_2}{\hat{s}_{p_1-p_2}}$$

在虚无假设下，它服从近似的标准正态分布。就新教徒数据而言，$z=0.143\ 2/0.040\ 9=3.503$，有双尾 $p$ 值【Value】为 0.000 46。这样，我们拒绝主流和原教旨新教徒赞成未婚女性流产相等的虚无假设。

在取平方时，这个 $z$ 统计量就变成了用于检验 2×2 列联表的关联的卡方【Chi-Square】($\chi^2$)检验统计量。对于新教徒数据而言，$\chi^2=z^2=12.27$，它与 $z$ 统计量相似，有 $p$ 值 0.000 46。

当所有的表计数 $n_{ij}$至少在 5 左右时，$z$ 和$\chi^2$ 检验被认为是合理近似的。对那些有低计数的表而言，应避免使用 $z$ 和$\chi^2$，建议使用费希尔的精确检验(Fisher's exact test)

(Hollander & Wolfe, 1999)。

——Paul T. von Hippel

(沈崇麟译校)

## 参考文献


Agresti, A., & Cato, B. (2000). Simple and effective confidence intervals for proportions and differences of proportions result from adding two successes and two failures. *American Statistician*, *54*(4), 280-288.

Agresti, A., & Finlay, B. (1997). *Statistical methods for the social sciences*. Upper Saddle River, NJ: Prentice Hall.

Hollander, M., & Wolfe, D. A. (1999). *Nonparametric statistical methods* (2nd ed.). New York: Wiley.


## 差分(Difference Scores)

差分是两个变量值之间的差别,通常都是在不同时点的同一变量上的两个量度值。例如,假如变量 $X$ 是一个侵略性态度量表,并在时间 1 和 6 个月之后的时间 2 用于调查对象。这样 $X_1-X_2$ 就等于侵略性从一个时间到下一个时间的差分或变化。有些变量随时间进展被反复量度,如一个国家的杀人犯数目会显示出一种上移的趋势。为了拉平这种趋势,诸如这样的变量可以转换成一种差分。

——Michael S. Lewis-Beck

(沈崇麟译校)

*也可参见变化分数【Change Scores】。

## 维度(Dimension)

这是保罗·拉扎斯菲尔德(Lazarsfeld, 1958)所使用的术语,用于指称一个概念【Concept】可以假定的不同方面。在概念可根据先验的理由被设定为不同的组成部分这个意义上,许多概念可看成是多维的。维度还可以指数据矩阵【Matrix】或列联表【Contingency Table】中的定向扩展,例如一行(向量或变量)、一列(向量或变量)等。

——Alan Bryman

(杨可译　沈崇麟校)

## 参考文献


Lazarsfeld, P. (1958). Evidence and inference in social research. *Daedalus*, *87*, 99-130.


## 解聚(Disaggregation)

解聚就是将聚集拆分成越来越小的单位。地理单位上的数据,比如说美国的州,可以拆分成国会选区或较小的静态单位县。聚集数据常常是将较小的地理单位聚集成

较大的地理单位而来,如上面的例子那样。不过它也可能涉及人的总体,例如在群体水平观察的行为,如一个工作组织中的合作问题就有可能解聚到个人的层次,譬如个体工作者的合作性。我们还可以再举另外一个例子,美国总统选举的投票行为,既可以解聚到个体投票者的层次进行研究,也可以将国家作为一个整体,在高度聚集的国家总统选举结果的层次上来加以研究。

——Michael S. Lewis-Beck

(沈崇麟译校)

## 话语分析(Discourse Analysis)

有很多理论和方法都采用了话语分析这个术语,主要是因为各种话语分析(DA)发源于不同的学科。所有的DA都关注话语或谈话和文本的结构及功能。不过,大多数DA彼此之间没有明显的关联。在这里,我们把关注的重点放在了那种与社会行动和互动紧密关联,且主要是在社会科学内部尤其是社会心理学中发展起来的DA。这就排除了“话语处理”的认知心理学,以及在语言学内超过单句界限的语法和语用分析的延伸发展。从特恩·迪杰克(Van Dijk's, 1985)的四卷本,以及本书其他一些条目,尤其是以语言学为根据的批判话语分析【Critical Discourse Analysis】,基于社会学的谈话分析【Conversation Analysis】,以及源自语言学和文学批评的叙事分析【Narrative Analysis】这些条目中,都可以领略各种各样分布广泛的话语分析。

乔纳森·波特和玛格丽特·韦瑟雷尔(Potter & Wetherell, 1987)在他们影响深远的著作《话语和社会心理学》(*Discourse and Social Psychology*)中,对那些主要用于各种事物的日常描述、评估或说明的话语,做出了理论界论,并明确了其研究路数。各种说明有三个主要特征:建构、功能和可变性。

①建构有两个方面的意思。话语既是建构的(用现成的语言资源来构造的),也是建设性的(提供这个世界的各种版本,作为不同于其他备择版本的特殊版本)。

②话语总是述行的,或者功能的。换言之,日常谈话和文本中说到或写到的东西,总是无一例外地产生于某种互动或争论的情境之中,在这种情境中,人们进行诸如评价、批评、请求、忏悔、诉求、辩护、拒绝等这样的行动。

③对事物的说明都是可变的,也就是说我们不应该指望人们总是前后一致。正是可变性这一性质提出了对传统的理论和方法的深刻批判,既包括对态度的性质和测量的批判,也包括对访谈实施和解释的批判。

这三个特征之间是相互关联的。无论描述的是什么,话语都是建设性的,所以无限多数的备择描述总是可能的,只要这些描述未简单或客观地变成了错误。这允许任何特殊的描述都是功能的或者表述行为的(action-performative),这不只是在诸如要求或问题这样的使用一般性文字的情况中,而且在外表简单、直白的事实描述中也同样如此。DA关心的主要问题确实一直都是事实描述所做的建构和功能工作。不仅如此,考虑到各种版本实现的功能工作以及跨版本的

可变性，即使是同一个说话人的言谈，也基本可以预期，它们都是以不同的版本在不同的场合做不同的事。这一原则也再一次产生了对其他方法的意义深远的批评，包括对使用访谈【Interview】和问卷【Questionnaire】的传统方法的批评，因为传统方法为了将个人的态度或理解定义成前后一致的，可变性被系统地避免或消除了。

## 历史发展

话语分析的起源包括语言哲学、后结构主义【Poststructuralism】、修辞学【Rhetoric】、常人方法论【Ethnomethodology】。这些起源一起合并成为一种话语方法。该方法专注于话语在社会实践中的位置及其在定义、合法化和破坏这个世界的事实版本中的作用。DA 直接的起源是科学社会学。奈杰尔·吉尔伯特和米歇尔·马尔基（Gilbert & Mulkay, 1984）发现，科学家在访谈、非正式的谈话以及在科学论文中给出的作为做出和反对知识宣称之根据的论述存在着不一致的问题。有两个重要的分析步骤是必需的。第一，他们必须决定把所有的有关真理和谬误的观念都当作是参与者的构建，并专注于科学家是如何制造它们的，而不是将当前的科学共识作为他们自己判断对错的标准。第二，他们要选定去制造诸如不一致性或变异性这样的中心现象，而不是去考虑如何去除不一致性。在科学论述中形成了两个对比鲜明的阐释套路【Interpretative Repertoires】：一个解释科学事实的经验主义套路和一个用个人和社会因素及缺点来解释错误的偶然套路。这项研究强调，是这两种解释套路，而不仅仅是标准的经验套路才是科学解释的基本要素。

波特和韦瑟雷尔的贡献是把科学社会学中的这些发展拓展成了一种在一般意义上分析话语的广泛应用的理论和方法。他们先是把建构、功能和可变性三个原则以及解释套路加以扩展，变成了对社会心理学中的态度理论与测量的研究和批评。波特和韦瑟雷尔没有去寻找诸如“态度”这样的假定在人们所说的话中表达的前后连贯一致的心理状态，而是和迈克尔·比利希等其他以话语为基础的批评家一起，去关注在言谈中生成的描述和评价的方式。话语的可变、功能以及表述行为的性质又一次崭露头角，并再次产生对较早建立的方法的批评。在使用这些早期的方法时，为了追求一致性、可靠性或测量集中趋势，通常都会将这些东西从视线中隐去。

最近的 DA 研究有两个主要方向，反映了不同的方法重点和研究主题。一个方向是把“解释套路”发展成一个能把话语实践的细节，一般来说是访谈谈话的转录文字，与更大的历史和意识形态主题联系起来的概念。套路被认为提供了供人们在构建关于这个世界的人和事的观念时提取的资源。这一派 DA（Wetherell & Potter, 1992）重点关注有清晰的意识形态意涵的有争议的主题，例如种族、性别、歧视行为等。它受福柯的话语概念影响，福柯认为，话语是历史演变而来的概念和解释框架，它与诸如法律、教育、医药、精神病治疗等体系的发展有关，应追溯话语对常识中的语言和意识形态的渗透。与福柯的话语不同的是，解释套路在分析上更易于处理，它更接近话语实践，也即由真正的言谈和文本实例所表现的行动和互动。DA 的第二个发展方向被称作话语心理学，其主要的分析动力源自谈话分析，主要关注对日常记录、自然发生的谈话的细致分析，也关注对文本资料如报纸报道的分

析,只是使用的范围要小一些。

## 话语心理学

话语心理学(Discursive Psychology, DP)是话语分析原则在心理学主题和概念上的应用(Edwards, 1997; Edwards & Potter, 1992)。尽管它的重点在心理学上,但其关联和应用的方位都很广泛。这是因为心理学的主题在用常识进行的解释中很普遍。日常言谈和很多制度性情境中关心的一个主要问题是人们的所感、所思、所解、所求、所欲和所虑的东西是什么,以及人们的个人禀赋等。事实上,这些主题通常都界定了制度环境下的标准的事务特性,例如,在教育环境、法庭、心理治疗和咨询等环境中,人们思考、理解和记忆的内容就是关注的要点。

DP有三个主要的流派:

①对心理学主题和解释的重构和批判,主要是把诸如"记忆""情感"等概念改造成话语实践;

②探究日常的心理学术语("心理词库")是如何使用的;

③研究言谈和文本中的心理学主题是如何不贴明显的标签来被处理和管理的。

这里使用心理学这个术语,并不是说这些主题和话题的处理是源于心理学的过程和解释。恰恰相反,DP要避免的正是支持社会行动语用学的心理学理论化。所有这三个流派似乎都与强调日常和制度性言谈和文本中的经验分析的实际研究有关。

## 分析原则

在DP中有一系列指导经验工作的分析原则。下面是几条代表性的而非限定性或排他性的原则。

①应优先考虑的是自然发生的数据或作为日常或制度性生活的一部分而产生的话语,而非通过调查(如研究访谈)得到的数据。虽然这不是问题的关键所在,但它反映了话语表述社会行动的原则;而许多社会科学家的"定性"数据都是访谈的对话,产生于讨论的生活之外,且只是反映了一点那些生活。

②不要问谈话/文本表达的心理状态是什么,也不要问它反映的世界状况是什么,而要问事情以那样的方式述说会有什么行动。

③要寻找参与者的顾虑,他们的类别、概念,以及他们正在处理的事情。探究他们是如何使用心理学概念或打消心理顾虑的。DP在任何阶段都不是为了要寻找潜在心理过程的证据。

④对任何分析者可能会带入数据的任何议题,试着对它做话题化的替代。也就是说,试着了解它在多大程度上是参与者本身在以某种方式操作或处理某件事情。

⑤关注主—客体关系(或者"心灵—世界"关系)。去寻找人们和他们的精神状态的描述是如何与行动、事件和外部世界的客体的描述相联系或为其所蕴含的,或者进行与之方向相反的探寻。

⑥去考察当下的讲话人/作者是如何反身性地看待他们自己的主客体问题的:他或她了解和讲述事物的根据;他或她是如何处理可能不为人所信、被人视为有偏或被认为掺杂情绪的。

⑦诚如所言,对于任何谈话内容,要问怎么,而不是为什么这样。要问"它是做什么的?它是怎么做到的?""为什么"问

题常常都取决于有关精神、语言、社会环境的规范性假设，因而一般最好把它们表达为某种“怎么样”的问题，最好不要问“为什么X那样说?”这样的问题，而要问“X那样说是不是可能有某些动机或理由?”这样的问题。

⑧要从修辞学上分析。要问“以这种方式讲话，否认了什么，反对了什么，预先阻止了什么……”

⑨要从符号学上来分析。就DA来说(也请参见符号学【Semiotics】)，这意味着去问“那些在这儿没有说的东西，是不是可以用相近的词汇或短语来说?”其原则是语言是一个有差别的系统，因此所有的词汇，所有的细节都有意义，因为有可供选择的(词汇或短语)。某个特定的词语或短语的选择是关键，因为通过想象那些看似合理的可供选择的(词汇或短语)和通过留心查看可供选择的描述的实际作用，我们可对其进行分析性的理解。这是一种考察所说的内容的特殊性的方法，而不是宣称讲话者真的喜欢哪个诸如这样的备择选择。

⑩要按顺序分析。对任何一段话语，要查看衔接在其前后的话语或者谈话的轮次，看一看现在的轮次或片段的内容处理和关联的是什么。这是谈话分析的主要原则，它庄严地载入了校验步骤【Proof Procedure】，而且出于同样的原因，它也是DP的一个基本要素。我们分析的不是被讲话者放入词汇中的思想的集合(像语录那样)，而是按顺序相关的方式进行的一系列活动。

⑪在发现复发模式时，要去寻找那些异常个案，它们是些看似不符合发展分析的个案，要查看分析是否需要改变，现象是否需要重新定义。不仅如此，异常个案分析【Deviant Case Analysis】还是一个源自会话分析的原则。

## 一个实例

DA要处理数量众多且种类繁多的数据，它们都不像在内容分析【Content Analysis】中的数据那样经过了编码和归类，而是像谈话分析中的数据，保持着全文本的、记录的或转录的形式。这就使得话语分析的所有原则都不太容易用一些简单的例子来说明。不过，我们却可以举例说明这种数据和涉及的分析。

下面的简单摘录来自一个婚姻咨询谈话，丈夫杰夫和妻子玛丽正在跟一位咨询师谈论他们婚姻中的困难。与这段摘录有关的是，玛丽有一段婚外情，她现在正在讲的是她当初在将这件事情告知杰夫时的反应(更多的细节可参见Edwards，1997)。

1.玛丽：嗯(1.0)，嗯(.)很明显。

2.你变得很生气。

3.是不是?

4.(.)。

5.看得出来很沮丧，唔，唔(0.6)。

6.我们开始吵个不停，嗯(0.6)。

7.只是疏……疏远了。

玛丽形容杰夫的反应是“生气”(第2行)、“非常沮丧”(第5行)。根据建构、修辞和语义的原则，请注意这些类别把杰夫的反应描绘成生气，而不是，比如说，已经要对他太太的行为和品德做出谴责的判断。他们把他的反应定义为情绪性的且是一种特殊的情绪。“生气”建立了各种可能性，这些可能性是认知“判断”，例如惩罚性判断所不具有的。情绪话语常常会利用其常识的不合理性，例如用不合理性将我们从对玛丽的推断引开，转而更多地去注意杰夫的精

神状态。同样的,一种看法、观点或判断可以(合乎规范地)比生气等情绪反应更持久,且很少变化。这是这一分析在此时此刻的概念以及修辞上的可能性。其重要性还取决于玛丽和杰夫接下来要说什么,因为这正是我们在试图寻找的——对他们的意义。

这个"你的愤怒阶段"的描述开始把生气的概念当作是一个有它的适当的场合和持续时间的临时状态来使用。这种描述蕴含了一种情绪反应的脚本(一个例行的,意料之中的事件系列)。例如,当确认杰夫的愤怒是合乎规矩的意料之中的(反应)时,那么一个愤怒阶段的想法便表达了一种意涵——它的持续时间不会长得离谱。玛丽在这里留下了修辞的空间,在这段摘录之后她也的确继续在此基础上发展出一个想法,就是杰夫的反应开始成了他们关系的一个问题。他的持续的情绪反应开始变成了咨询关注的焦点。咨询关注的焦点成了他的不恰当的感受,而不是她的不忠的行为和可能的不良倾向。

因此,"你愤怒的阶段"这个表达是建构的、修辞学的和述行的。它并不只是在描述杰夫(作为世界上的一个客体),也不只是表达玛丽思想上对杰夫的理解(把她的想法诉诸语言)。相反,它的作用是将他们之间的问题定义为一个存在于杰夫身上的问题,而与此同时,玛丽却在避免对杰夫的公然羞辱。确实,"你的"愤怒一词和合乎规范的概念"阶段"一起,以及"很明显"(第1行)这一说法,把他的反应描述成了一种规范授权的反应,至少在一段时期内如此。

也请注意这个描述是如何符合叙述的顺序的。玛丽下面要叙述的事情(而且,不露声色的是,不仅仅是跟着,而且是紧跟在杰夫的反应之后;参见有关"叙事"的词条)就是我们如何"开始吵个不停,就逐渐疏远了"(第6~7行)。他们的问题变成了双方的问题,成了一种没有过失方、没有受指责方,"只是"双方变得逐渐疏远了的问题。杰夫的"阶段"已经是过去时(你经过了你愤怒的阶段,对吗?),这里又一次暗指他的怒气应该消了。通过具体的细节,玛丽所生成的讲述总体方向是让我们在理解他们之间的问题时,把重点从她的婚外情转移到她丈夫在面对这件事时持续的情感困扰之上。

——Derek Edwards

(杨可译　沈崇麟校)

## 参考文献


Edwards, D. (1997). *Discourse and cognition*. London: Sage. Edwards, D., & Potter, J. (1992). *Discursive psychology*. London: Sage.

Gilbert, G. N., & Mulkay, M. (1984). *Opening Pandora's box: A sociological analysis of scientists' discourse*. Cambridge, UK: Cambridge University Press.

Potter, J., & Wetherell, M. (1987). *Discourse and social psychology: Beyond attitudes and behaviour*. London: Sage.

Van Dijk, T. A. (Ed.). (1985). *Handbook of discourse analysis: Vols. 1-4*. London: Academic Press.

Wetherell, M., & Potter, J. (1992). *Mapping the language of racism: Discourse and the legitimation of exploitation*. Brighton, UK: Harvester Wheatsheaf.

## 离散(Discrete)

这一术语特指一种诸如分类【Categorical】变量这样的变量,这种变量将单位编组,成为离散的集群。它也常常用来指这样一些事例,在这些事例中组成变量的单位是离散的。因此,孩子数作为一个变量只能有离散的单位——我们不可能有3.67个孩子。

——Alan Bryman

(沈崇麟译校)

* 也可参见类型【Attribute】、分类的【Categorical】、定类变量【Nominal Variable】。

## 判别分析(Discriminant Analysis)

社会学家弗瑞德·普鲁沃(Provoost,1979)在一个研究中对判别分析的基本思想作了清楚的阐释。他在比利时讲荷兰语的地区和布鲁塞尔地区进行了一次贫富居民区的差别的科学调查。他区分出了六个判别变量[平均收入(mean income)、受教育程度(educational attainment)、职业(occupation)、住房状况(housing situation)、社会和文化参与指数(an index of social and cultural participation)及服务设施(the presence of services)],并将这六个变量合在一起来考察它们显著判别这两种居民区的能力。

这项涉及两个社会群体判别的研究由三个步骤组成。第一步,要尽可能可靠地对群体预先进行分类。因为这个分类正是我们要解释的题目,它代表(二分的)有着诸如贫民区这样的分类的因变量。第二步,收集许多特征的统计数据,这些特征的判别能力取决于使用的统计分析方法,同时还要计算加权的总和。这些特征就是自变量$X_1$,$X_2$,…(判别变量),它们的权重(判别权重)分别为$k_1$,$k_2$,…;加权总和是$f = k_1X_1 + k_2X_2 + \cdots$。第三,如果第二步的分析似乎是成功的(加权总和与每个单独的特征都有显著的判别能力),那么我们便有了一个能将其他单位分类并归入这两种群体中的某一个的精良工具。最后一步,分类,给我们提供了一个有根据地确定那些未曾包含在分析中的居民区究竟是富还是穷的机会。

延续前述,判别分析的结构除了判别分析的因变量是分类的(在本例中,甚至是二分的),而不是定距的之外,其余都与多元回归【Regression】分析相同。图1是阐释研究问题的因果图。

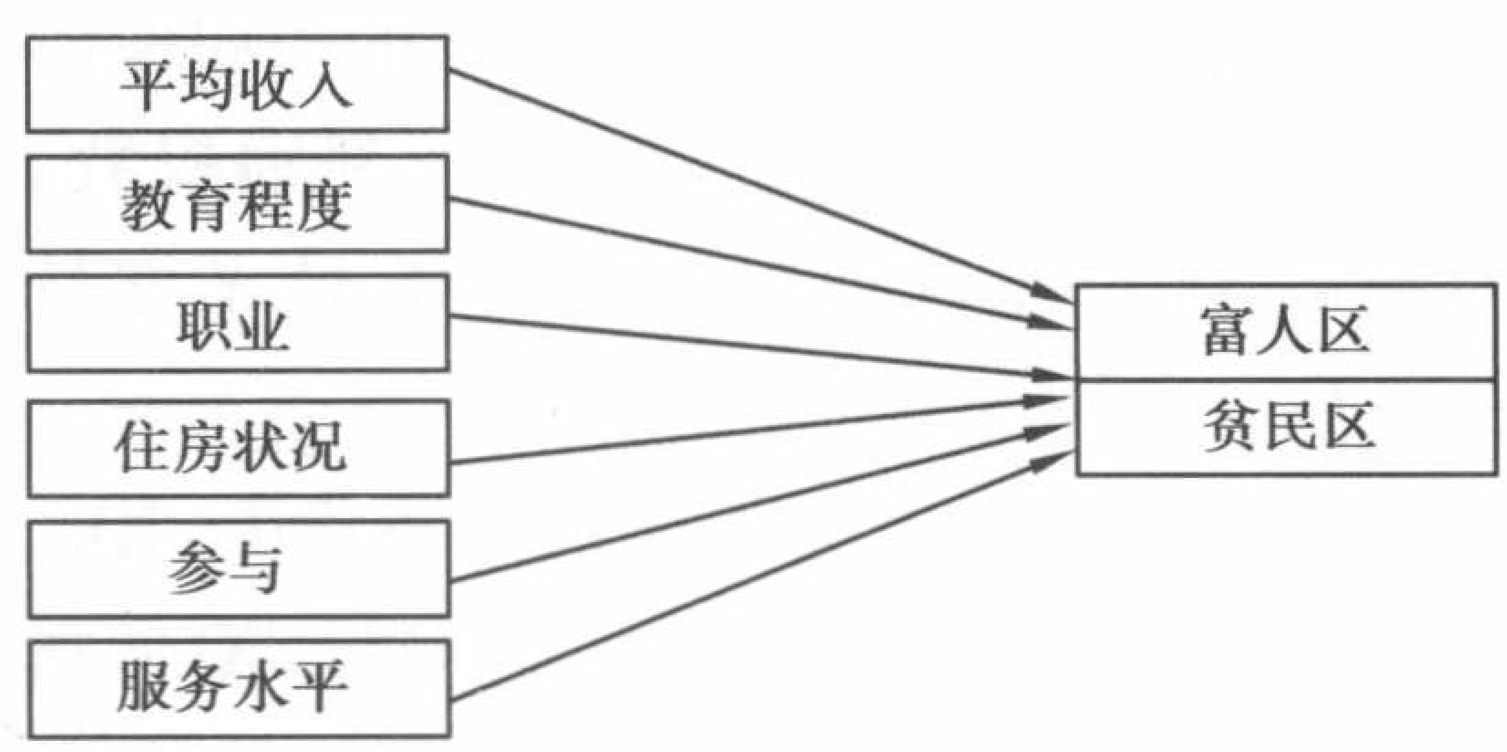



图1

分析单位是地区或居民区。判别变量平均收入($X_1$)、教育程度($X_2$)……服务水平($X_6$)是在定距或定比层次量度的。与多元回归分析一样,它们可看成是多因模型中的起因。因变量$Y$代表群体。它可被看成是一个分类变量,而在这里则是一个有贫、富两类的二分变量。考虑到也许会使用虚

拟变量的分析方法，我们可以对它们分别采用0和1的编码。我们也将不再预测 $Y$ 的得分，而是要将它分入两个群体中的某一个。与多元回归分析一样，这个模型也是可加的。这意味着，原则上不存在交互效应，即我们要考虑的是加权的 $X$ 变量之和，而不是像 $X_1X_2$ 这样的乘积项。

## 判别分析模型

在判别分析中，我们想要考察一组变量 $(X_1, X_2, \cdots)$ 是否能在一些群体之间作出区别(即在中性意义上判别)。因此，我们将以群体(贫民区和富人区)做出最大区别的方式来寻找判别变量 $(X_1, X_2, \cdots)$ 的线性组合。这样的线性组合称为判别函数，而它们一般都具有这样的形式

$$F - \overline{F} = k_1(X_1 - \overline{X}_1) + k_2(X_2 - \overline{X}_2) + \cdots + k_p(X_p - \overline{X}_p)$$

或

$$f = k_1x_1 + k_2x_2 + \cdots + k_px_p$$

在这个公式中，$f$ 和 $x_i$ 被表示为均值的偏差(小写字母)。系数 $k_i$ 称为判别权数。变量 $x_1$ 到 $x_p$ 是判别变量，其中 $p$ 是 $X$ 变量的总数。

这个公式可以更为简单地写成 $\mathbf{f} = \mathbf{Xk}$，其中的 $\mathbf{f}$ 是判别分数的 $n\times1$ 列向量，$\mathbf{X}$ 是判别变量分数的 $n\times p$ 矩阵，$\mathbf{k}$ 是判别权数的 $p\times1$ 列的向量。

在两个群体的情形中，只有一个判别函数。而在其他情形中，如果要比较的群体有若干个，那么判别函数的数目最大将等于 $\min(g-1, p)$。例如，如果用 $p=5$ 个判别变量的均值分析 $g=4$ 个群体，那么有 $g-1=3$ 和 $p=5$，这样，最大的判别函数数目就等于3。

## 判别分析的目的

判别分析的目的有三个。将贫富人区(两个群体)这一例子记在心中，并将自己限制在两个判别变量($x_1$ 和 $x_2$)内，那么这些目的可以被概括为以下几个方面：

1.我们要寻找一个称为判别函数的线性函数 $f = k_1x_1 + k_2x_2$。寻找这种函数的方法是，在这种函数 $f$ 上的居民区分数(判别分数)要具有最大化组间和组内的变差率这一性能。这可归结为判别权数 $k_1$ 和 $k_2$ 的计算和随后的 $n$ 个判别分数的计算。

2.我们要考察由函数 $f$ 建立的判别能否推广到总体。这两个群的型心之间是否有显著的差别？也就是说，综观变量 $X_1$ 和 $X_2$，两个群体的均值之间是否存在显著差别？这可归结于做一个 Hotelling $T^2$ 检验。这种检验是简单的学生 $t$ 检验的扩展。在更广的三个或更多的群体情形中，这便要归结为 Wilks 的 $\Lambda$ 检验。

3.使用判别函数 $f$。我们也希望确定每个居民区，进而再逐渐确定新的居民区属于(贫富)两个群体中的哪一个。假如我们已经知道了居民区的收入和服务水平的分数，我们能否预测它是富有还是贫穷？这个问题可归结为确定一个分界点(在两个群体的型心的中间)和验证判别分数究竟位于这一点的左边还是右边。如果一个判别分数等于分界点，那么相应的居民区便可被随机地分配给两个群体中的其中一个。

这种将单位分组的过程大多数都在统计决策论和贝叶斯统计这一大背景下考虑。

## 判别函数 $f$ 的计算

函数 $f$ 是 $x$ 变量的线性组合：$f = k_1x_1 + k_2x_2 + \cdots + k_px_p$。这个函数 $f$ 将以这样的方式计算，即它将最大可能地对两个群体作出判别，使组间和组内的离差的比率最大。为了

达到这一目的，$f$ 的变差将被分为组间和组内两个部分，而判别权数将以组间/组内比尽可能大的方式计算。

为了求解这种最大化问题，我们将要对典型矩阵的本征结构进行一番考察。在判别分析中，这一典型矩阵是 $\mathbf{W}^{-1}\mathbf{B}$，它表示变差和协变差的组间/组内比。这一词条末尾列出的大多数参考文献都详细介绍了本征结构的考察问题。

## 分类和预测

在找到判别函数之后，每个居民区期望的 $f$ 值（=判别分数），都可通过将 $X$ 变量的分数代入函数 $f$ 来计算。为了求 $f$ 分数，群体 1 的型心 $\overline{f_1}$ 和群体 2 的型心 $\overline{f_2}$ 我们也可作同样的运算。在 $f$ 轴上的这两个群的型心的预测值的中间点可选作分界点。如果两个群是等容量的，那么这个分界点将位于原点，即 $f_c=0$。在群体 1 含的居民区数少于群体 2 时，分界点将位于原点的左侧。

根据居民区的判别分，我们把每一个居民区分入两个群体中的其中一个。那些得分在 $f_c$ 右侧（$f_i>f_c$）的居民区被分入群体 2，而那些得分在 $f_c$ 左侧（$f_i<f_c$）的居民区则被分入群体 1。然后我们可以指明预测的群体成员资格。将这个与原来的 $Y$ 分数作比较，我们可以了解有多大百分比的居民区（如 80%）的分类是正确的。在两个群体之间发生重迭时，判别函数就无法完美地进行分类。

## 显著性检验

我们想知道两个群体的型心，也即综观 $X$ 变量，两个群体的均值之间是否有着显著的差别。这就是一个全模型检验。此外，我们也想分别进行显著性检验，即所谓的单变量检验，因为 $X$ 变量是一次只考虑一个的。

对于多变量检验而言，我们可以使用 Wilks 的朗姆达，即 $\Lambda$，就像学生 $t$ 和费希尔的 $F$ 一样，它要将样本的组间离差和组内离差比率与虚无假设下的这一比率作比较。Wilks 的 $\Lambda$ 是它的最为一般形式中的这一比率的多变量推类（multivariate analogy），所以 $t$ 和 $F$ 是 $\Lambda$ 的特例。因此，这一量度被用于更为复杂的多变量分析技术，如方差【Variance】和协方差【Covariance】的多元分析【Multivariate Analysis】、典型相关分析【Canonical Correlation Analysis】和多重判别分析。

在问题只涉及两个群体的时候，同本例一样，可进行 Hotelling 的 $T^2$ 检验，它是朗姆达检验的一个特例。$T^2$ 有一个性能，即它包含了一个特殊的矩阵 $d'C_w^{-1}d$，称为 Mahalanobis 距离。在这一广义的距离中，将不等的判别变量的离差和它们的相互关联也都考虑了进去。

## 扩展

判别权数虽然一般都用数据的本征结构来计算。但其他公式也是可以使用的，例如从总样本协方差或相关矩阵（$\mathbf{C}_T$ 或 $\mathbf{R}_T$）开始，或用虚拟变量 $Y$ 对判别变量的回归得到的虚拟回归法，最终用比例转换得到相同的判别权数。总之，判别权数的计算方法有很多选择，使我们有无从着手的感觉，以致会面临保罗·格林的"财富的窘境"。

在前述讨论中，变量都假定是在定距或定比层次量度的。同时它也假定关系都是线性的。不过这两个要求是可以放松的。在 SPSS 之外，来自荷兰莱顿大学的计算机软件 CRIMINALS 也可用来作判别分析，但它是非度量（nonmetric）和非线性的。有兴趣的读者可参阅吉非的有关著作（Gifi，1980）。

线性判别分析也是一种有许多额外的要求的方法，除了线性之外，还有常态性和协方差矩阵的同次性也必须得到满足。在协方差矩阵确实显著不同时，二次判别函数可能更

为合适。有关这一专题的文献可参阅拉亨布鲁克的著作(Lachenbruch, 1975)。在严重违反判别分析应满足的要求时,logistic 回归【Logistic Regression】也许总是不失为一种很好的替代方法。

——Jacques Tacq

(沈崇麟译校)

## 参考文献


Cooley, W., & Lohnes, P. (1971). *Multivariate data analysis*. New York: Wiley.

Gifi, A. (1980). *Nonlinear multivariate analysis*. Leiden, The Netherlands: Leiden University Press.

Green, P. (1978). *Analyzing multivariate data*. Hinsdale, IL: Dryden.

Lachenbruch, P. A. (1975). *Discriminant analysis*. New York: Hafner.

Tacq, J. J. A. (1997). *Multivariate analysis techniques in social science research: From problem to analysis*. London: Sage.

Van de Geer, J. (1971). *Introduction to multivariate analysis for the social sciences*. San Francisco: Freeman.


# 判别效度(Discriminant Validity)

判别(也称为发散)效度是一种量度没有过分关联于其他类似但有区别的构造的证据(Messick, 1989)。一个构造的量度和概念上不同的构造的量度之间的相关【Correlation】系数通常都被作为判别效度的证据。依据理论关系和系数的数量,一方面,如果相关系数高,这就表明判别效度的缺乏或微弱;另一方面,如果相关系数低或中等,这就证明量度有判别效度。判别效度尤其与无关变量可能对分数有影响的检验确认有关,例如,证明一种激情之爱的量度与社会赞许无关这一点是很重要的。

证明判别效度的其他方法是多质多法(MTMM)(参见混合法研究【Mixed Methods Research】)和因子分析【Factor Analysis】法。MTMM 对考察收敛和判别效度二者都很有用。例如,这种程序用一种以上的方法量度一个以上的构造。如一种新的工作满意度量度,可能与现存的生活满意度和关系满意度的有效量度相关。我们可以得到自我报告的数据,以及督导和合作者在三种构造中的每一种上的打分(也即三种不同的方法)。在理论上,工作、生活和关系满意度是在概念上不同的三种构造,因此,为了显示判别效度,这些量度之间的相关程度应该比较低。同样,如果判别效度存在,那么来自不同的观察者的不同的构造的打分之间的相关也应该比较低。

因子分析可以用于显示相似的但概念上不同的构造是有区别的;例如皮尔斯、加德纳、卡明斯和邓纳姆(Pierce, Gardner, Cummings, & Dunham, 1989)用若干情感性量度(如组织承诺)的因子分析来评估他们新开发的基于组织的自尊(organization-based self-esteem)量度。他们的发现,显示基于组织的自尊是可以与情感量度区分的,因为与其他量度比较,它们负载于不同的因子。相反,因子分析也可用于缺乏判别效度的证明。卡尔斯(Carless, 1998)用验证性因子分析证明了一个变革型领导量度的子量表彼此之间不仅高度相关,而且形成了一个变革型领导的高阶构造。在研究概念上重叠的变量和收集的数据来自一个来源(如一个个别的回答人)时,因子分析是一种基本的分析技术。例如,忧郁和焦虑是在理论上有区别的构造,但在经验上,它们之间也

存在着重叠，忧郁的个体也经常高度焦虑。为了验证量表，重要的是我们不仅要考查变量之间的相关的数量，也要用因子分析证明判别效度。

——Sally A. Carless

（沈崇麟译校）

## 参考文献


Carless, S. A. (1998). Assessing the discriminant validity of the transformational leadership behaviour as measured by the MLQ. *Journal of Occupational and Organizational Psychology*, *71*, 353-358.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13-103). New York: American Council on Education and Macmillan.

Pierce, J. L., Gardner, D. G., Cummings, L. L., & Dunham, R. B. (1989). Organization-based self-esteem: Construct definition, measurement, and validation. *Academy of Management Journal*, *32*, 622-648.


# 反序交互(Disordinality)

反序交互是一种交互【Interaction】作用，在这种作用中，$X_1$ 和 $Y$ 之间的关系【Relationship】依据 $X_2$ 的值而不同和相反。在一般的交互作用中，依据 $X_2$ 的值，$X_1$ 对 $Y$ 的影响只是有所不同而已。假设我们的问题是美国劳动力人口样本中的男性和女性（$X_2$，1 = 男性，0 = 女性）的教育（$X_1$，以年量度）与收入（$Y$，以美元计）的关系。我们现在来做 $Y$ 在 $X$ 上的回归，先做男性的子样本，再做女性的子样本。如果我们的发现是男性的回归【Regression】斜率明显比女性的更陡峭，那么我们可以得出存在交互作用的结论，因为教育对收入的影响取决于性别。这一结果就是一个普通的交互作用。如果与此不同，我们发现男性的斜率是正的，而女性的斜率不仅是显著不同的，而且还是负的，那么这种交互作用就是一种无序的交互作用。在这种交互作用中，在一个群体内，$X_i$ 有着正效应，而在另一个群体内，$X$ 则有着负效应。在这一例子中，两根回归线不仅像普通的回归那样，是不平行的，而且实际上还是互相交叉的。想象一下这样的情况，对问题的理解可能不无益处。

——Michael S. Lewis-Beck

（沈崇麟译校）

* 也可参见定序交互【Ordinal Interaction】。

# 离中趋势(Dispersion)

离中趋势大致上是指收集的一组观察的离散度或变异性。例如，个人关于政治领导人的业绩这一问题的回答可能会不尽相同，孩子们对一种旨在增强阅读能力的方法的反应可能也会有所不同，即使在物理科学中，一次和另一次量度，因为使用的工具的精确度问题，也可能会有差别。从某种意义上讲，引起我们对于统计技术的兴趣的恰恰就是离散趋势问题。

一个基本的问题是在我们试图描述一个个体或物体的总体【Population】特征时，应当确定如何来测量离中趋势。也就是说，如果所有感兴趣的个体是可以量度的，那么在这些个体中间的变差应该是什么样的？

这样的量度是离中趋势的总体量度。一个与之关联的问题是怎样依据个体的样本【Sample】来估计离中趋势的总体量度。选择离中趋势的量度是一个复杂的问题,它在过去的30多年中有了长足的进步。这种选择在一定程度上取决于调查研究者的目的,因为最佳选择常常会随调查研究者的所思和所为而剧烈地变动。建议采用的离中趋势量度有150多个,根据与一系列实际问题有关的基本标准,劳克斯(Lax, 1985)对它们作了比较。虽然这些量度中的大多数并没有什么实际的价值,但至少有5~6种是重要和有用的。

不言而喻,最著名的离中趋势量度是总体方差【Variance】,它一般写作$\sigma^2$。它是观察和总体均值【Mean】$\mu$之间的平方差的平均(或期望)值。那就是说,如果所有的个体都是可以量度的,那么他们的回应的平均值称为总体均值$\mu$,而如果对每个观察而言,它和$\mu$之间的平方差是可以计算的话,那么,这些平方值的平均值就是$\sigma^2$。在更为正规的表述中,$\sigma^2 = E(X-\mu)^2$,其中$X$是我们可能做的任何观察,而$E$则表示期望值。(正的)$\sigma^2$的平方根$\sigma$称为(总体的)标准差【Standard Deviation】。依据一个$n$个个体的样本,如果我们观察值$X_1, \cdots, X_n$,那么$\sigma^2$的估计值通常就是样本方差:

$$s^2 = \frac{1}{n-1}\sum(X_i - \overline{X})^2$$

式中的$\overline{X} = \sum X_i/n$是样本均值。

对某些目的而言,标准差$\sigma$的使用源自这样一个基本结果,即一个观察值在位于离均值的某个以$\sigma$量度的特定的距离之内的概率。这种概率在正态性下是完全确定的。例如,一个观察值在均值的一个标准差之内的概率是0.68,而两个标准差之内的概率是0.954。这些性质产生了一个普遍使用的效应量【Effect Size】量度(一个描述两个群体差异程度的量度),以及经常被用来确定特异值【Outlier】(一般是很大或很小的值)的法则。在图基(Tukey, 1960)影响深远的论文发表之后不久,人们认识到,即使是对正态性很小的偏离,也可能使这些性质有实质性的改变,从而经常会导致实际问题发生。

例如,我们来考虑一下图1显示的两个分布:一个是标准正态分布,这意味着它有均值0和方差1。另一个则不是正态分布,或许称为某种混合或被污染的正态分布。这就是说,两种正态分布【Normal Distribution】有着相同的均值,但却有不同的方差,被混在了一起[这种特别的混合的正态的精确描述请参见威尔克斯的著作(Wilcox, 2003),在这里我们只作一些简单的介绍]。问题的重点在于,尽管两个分布之间存在着相似之处,但它们的方差却存在实质性的差别。正态分布有方差1,但混合正态有方差10.9。这就阐明了一个普遍原理:分布尾部的微小变化,便会对$\sigma^2$产生实质性的影响。

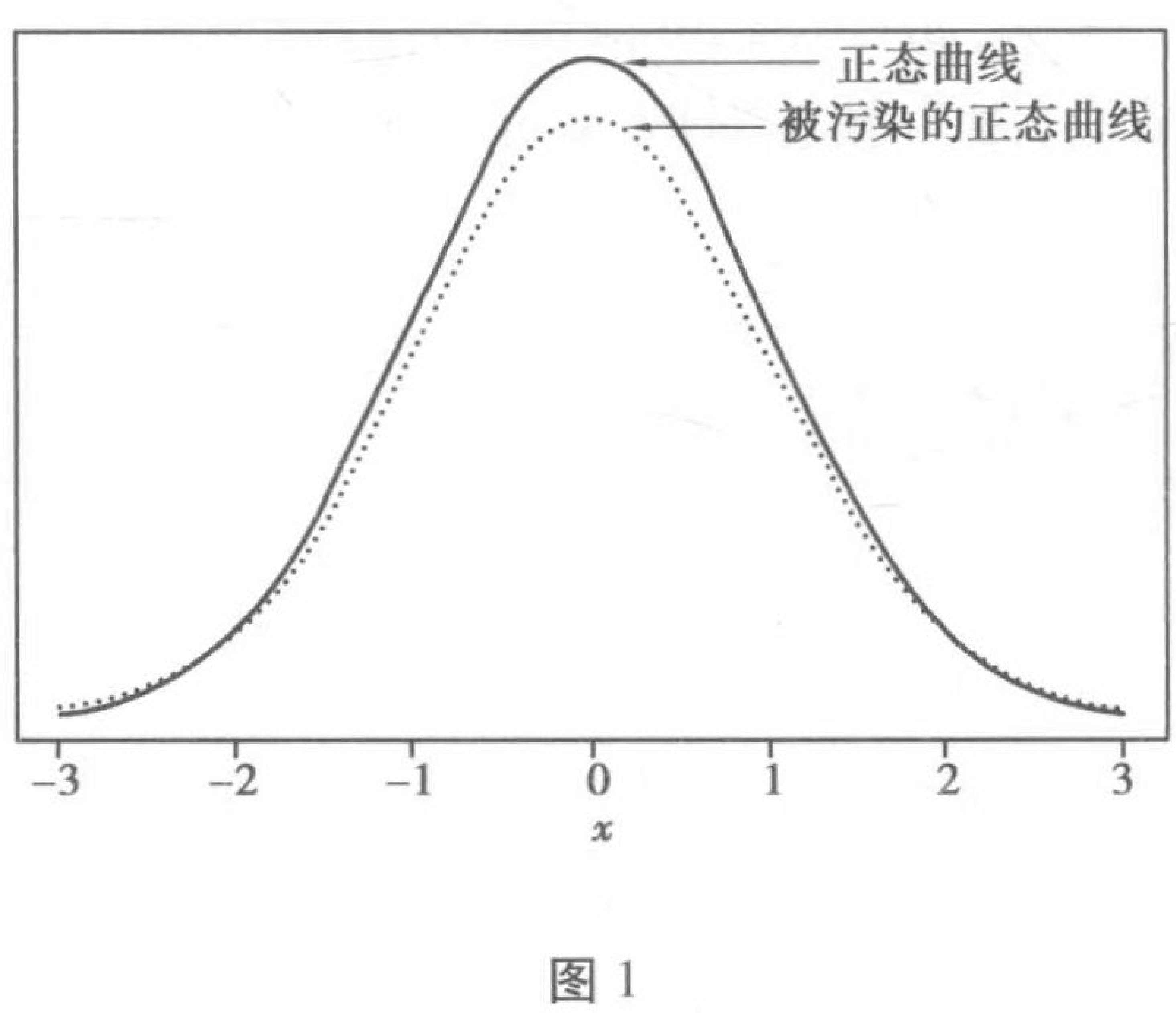



图1

现在考虑在正态性下的性质,一个观察位于离均值1个标准差之内的概率是0.68。

这一性质在非正态性下是否也存在？在有些情况中，这近似于真的，但在另外一些情形中，情况便将不再如此，即使是对正态性的些许偏离（如同统计学家用几种普遍使用的度量所量度的一样）。对于图 1 中的混合正态而言，这个概率超过了 0.999。

样本方差反映出同样的性质：它对特异值【Outlier】非常敏感。在有些情形中，这一性质非常有益处，但是对于一般性问题而言，它却是有害的。例如，一个普遍使用的法则是，如果一个值 $X$ 距均值两个以上（样本）标准差，我们就将它视为特异值，这一思想是以正态性为依据的，从概率的观点看这样一个值是异常的。这就是说，如果

$$\frac{X - \bar{X}}{s} > 2$$

我们就称 $X$ 是一个特异值。

然而，这一法则常常会被屏蔽，这就意味着，因为 $s$ 对特异值具有的敏感性，它甚至连明显的特异值也探测不出来。例如，对于 2，2，3，3，3，4，4，4，100 000，100 000 这样一组值，$\bar{X}=20\ 002.5$，$s=42\ 162.38$，而 $(100\ 000-20\ 002.5)/s = 1.9$。这样，根据刚才给出的法则，我们就不能把 100 000 确定为异常值，然而问题很清楚，对于其他的那些值而言，这个值显然是不寻常的。

为了探测特异值，一般可用两种备择的离中趋势量度。第一种备择量度称为绝对中位差（MAD）统计量。它是 $n$ 个值

$$|X_1 - M|, \cdots, |X_n - M|$$

式中的 $M$ 是 $X_1, \cdots, X_n$ 的平常的中位数。如果

$$\frac{X - M}{\mathrm{MAD}/0.674\ 5} > 2.24$$

我们便可把 $X$ 视为特异值。

第二种备择量度称为四分位间距【Interquartile Range】，它无非就是上四分位数和下四分位数的差（后一种量度为探测特异值的箱形图法则所用）。这两种离中趋势的量度的一个重要特点是，它们都对极端值不敏感，反映分布在中心的观察的离差。在做旨在做避免屏蔽的异常值探测时，这正是一种合乎要求的性质。

选择一个离中趋势量度的另一种方法是，寻找一个在相当广泛的情形中有一个比较低的标准误【Standard Error】的估计量【Estimator】。尤其是，它不仅在正态性下有一个有竞争力的 $s$，而且希望在非正态性下它也能保持一个较低的标准误。因此，如果它是一个理想的比较两个群体的离中趋势的（量度）话，那么无论是从一个正态分布，还是从图 1 中的混合正态分布抽样，它均能保持高效力。这大致与拉克斯（Lax，1985）论及的问题相当。在许多方法中，有两种估计量脱颖而出。第一种称为弯曲中方差百分比（percentage bend midvariance），第二种称为双权中方差（biweight midvariance）。这两个估计量的计算方法在这里不作介绍。对此问题和使用方面的应用软件有兴趣的读者，可参阅威尔克斯的有关著作（Wilcox，2003）。就达到比较低的标准误而言，样本标准差 $s$，在正态性下可与这两个量度比美，但在非正态性下，则 $s$ 可能就大为逊色了。

## 效应量

在其他各种情形中，方差起着重要的作用。这在一定程度上是因为它具有为正态分布所喜爱的特性。这些量度中，一种使用较为普遍的是描述两个群体之间的差异的

效应量量度,即

$$\Delta = \frac{\mu_1 - \mu_2}{\sigma_p}$$

式中的 $\mu_1$ 和 $\mu_2$ 是总体均值。而由假定可知,两个群体有相等的标准差 $\sigma_p$。它的基本思想是:例如,若 $\Delta = 1$,那么两个均值之间的差便是 1 个标准差,它显示了正态性下群体的差别有多大。但对于这种特殊的方法的担忧也随之而起,因为在非正态性下,它可能会屏蔽比较大的差别(Wilcox, 2003)。

## 标准误

变差的概念将以下面的方式扩展到抽样分布【Sampling Distributions】。设想一个基于 $n$ 个观察的研究,得到了一个样本均值,比如说是 $\overline{X}$。现在我们再设想,研究重复了许多次(在理论上是无限多次),得到样本均值 $\overline{X}_1, \overline{X}_2, \cdots$,而每个样本均值也都以 $n$ 个观察为根据。这些样本均值的方差称为样本均值的平方标准误,在随机样本下,已知它等于 $\sigma^2/n$。也就是说,它是样本均值的方差。这些方差 $E(\overline{X}-\mu)^2 = \sigma^2/n$。不言而喻,这个方差的最有用和重要的作用是根据样本均值推论【Inferences】总体均值。因为样本均值的标准误 $\sigma/\sqrt{n}$ 告诉我们如何在正态性假设下进行对 $\mu$ 的推论。但有关这方面的内容超出了本词条的范围,对此感兴趣的读者可自行参阅有关的著作。

## 缩尾方差

近年来,另一个显得越来越重要的、在基于截尾均值(trimmed means)的群体比较中起重要作用的离中趋势量度就是缩尾方差。它是按如下步骤进行计算的。考虑任何诸如 $0 \leqslant \gamma < 0.5$ 这样的 $\gamma$。令 $g = \gamma n$,并将它下舍到最接近的整数。这样,如果 $\gamma n = 9.8$,则 $g = 9$。计算一个 $\gamma$ 截尾均值则需将 $g$ 的最小和最大的观察值去除,然后再取其余的观察值的平均数(参见齐整【Trimming】)。对 $n$ 个观察进行缩尾处理,是要将 $g$ 的最小值置为未经齐整时的最小值,同时将 $g$ 的最大值置为未做齐整时的最大值。基于缩尾值的样本方差称为缩尾方差。理论和仿真研究二者都表明,齐整可以减少关乎因非正态性产生的与均值关联的问题,在样本量比较小时,情况尤其如此。尽管不是很直观,但理论表明,齐整均值的标准误差的平方与缩尾方差有关(Staudte & Sheather, 1990),因而在检验基于齐整均值的假设时,缩尾方差起着很重要的作用。不仅如此,在寻找类似皮尔逊相关【Pearson's Correlation】的稳健的类似统计量时,它同样也是很有用的(Wilcox, 2003)。

还有许多被建议使用的多变量离中趋势量度,其中若干个现在仍有相当大的实用价值。而这些量度中的某几个在卢苏和勒罗伊(Rousseeuw & Leroy, 1987),以及威尔克斯(Wilcox, 2003)等人的著作中都有介绍。

——Rand R. Wilcox

(沈崇麟译校)

## 参考文献

Lax, D. A. (1985). Robust estimators of scale: Finite-sample performance in long-tailed symmetric distributions. *Journal of the American Statistical Association*, *80*, 736-741.

Rousseeuw, P. J., & Leroy, A. M. (1987). *Robust*

regression & outlier detection. New York: Wiley.
Staudte, R. G., & Sheather, S. J. (1990). *Robust estimation and testing*. New York: Wiley.
Tukey, J. W. (1960). A survey of sampling from contaminated normal distributions. In I. Olkin et al. (Eds.), *Contributions to probability and statistics*. Stanford, CA: Stanford University Press.
Wilcox, R. R. (2003). *Applying contemporary statistical techniques*. San Diego, CA: Academic Press.

## 相异(Dissimilarity)

相异是一种用来描述相似或相异的量度的通用术语(也称“接近”量度,它常与多维度量【Multidimensional Scaling】模型和聚类分析【Cluster Analysis】共同使用)。这样的量度可能是:(a)一种由被试者做出的直接的判断或相似性的等级打分;(b)或源自一变量的总计量度。直接的判断是用按从“完全相似”到“完全相异”排列的等级量表的两个对象之间做出的。然后这一判断可作为“准距离”和直接测量。总计量度随量度的层次不同而有所不同,且通常都是非相关的,因为对于距离的模型表示而言,一个相异量度不仅应该是非负和对称的,而且还应当是服从三角不等式的。

——Anthony P. M. Coxon

(沈崇麟译校)

### 参考文献

Gower, J. C., & Legendre, P. (1986). Metric and Euclidean properties of dissimilarity coefficients. *Journal of Classification*, *5*, 5-48.

## 分布(Distribution)

分布在社会科学中无所不在。它们出现在分析特定局部域(topical domains)的框架中,为我们提供多种多样的形状来表示基本数量的可能变异范围;它们出现在理论中,为我们提供推导预测的基本工具;它们也出现在要对潜在的分布形式作估计的经验工作中,并用分布来描述不可观察的运行特性。

### 分布的分类

分布可按几个维度进行分类——连续与离散、单变量与多变量、建模与抽样或按它们的连带函数的数学特性进行分类。有关这方面的导论和全面阐述,请参见约翰逊和考兹的系列丛书“统计分布(Distributions in Statistics)”及其修订本(Johnson, Kotz, & Balakrishnan, 1994),以及伊文思、黑斯廷斯和皮考科等人编撰的小手册(Evans, Hastings, & Peacock, 2000)。

### 连带函数

在数学上设定的分布有各种函数与它们联系。最为基本的是分布函数(又称为累积分布函数),它被定义为变量 $X$ 假定一个值小于或等于 $x$ 的概率 $\alpha$,一般表示为 $F_x(x)$ 或更为简单地表示为 $F(x)$。也许最为著名的连带函数是概率密度函数【Probability Density Function】,表示为 $f(x)$,

是一种连续分布,它是分布函数关于 $x$ 的一阶导数(是一种离散分布,有时也称为概率质量函数)。例如,在正态族中,钟形曲线【Bell-shaped Curve】描绘的概率密度函数是较卵形曲线描绘的概率分布更为普遍的表示法。对于社会科学而言,最为有用的两种函数都是分位函数,它们在其他各种用处之外,还给出了不平等的全分布量度的基础,如 Pen Parade[这是一个分配函数,它的意思是指,为了估测增长是否减少贫困,通常的做法是按平均收入高低的分布来计算平均收入,即"Pen Parode"(Pen,1971),其中一步就是计算平均收入最低的 1/5 人口的收入增长率——译者注]和风险函数(hazard function)。

当然,所有的连带函数都以某些特别的方式相互联系。如我们已经知道的一样,在连续分布中,概率密度函数是关于 $x$ 的分布函数的一阶导数。表示各异的分位数函数,$G(\alpha)$,$Q(\alpha)$ 或 $F^{-1}(\alpha)$ 是分布函数的倒数,提供了一种来自分位数【Quantile】$x$ 的概率的映像。这两种函数之间的一些重要关系包括(Eubank, 1988; Evans, Hastings, & Peacock, 2000):

$F[Q(\alpha)] = \alpha$,$F$ 连续的

$Q[F(x)] = x$,$F$ 连续的和严格递增的

## 分布参数

分布也有一些与之关联的参数。在所有这些参数中,有三个被视为是基本的——位置、尺度和形状参数。位置参数是变量域中的一个点,而尺度和形状参数则分别控制着量度的尺度和形状。有时跨分布族地使用一些统一的记号不失为一种明智之举。例如埃文斯、黑斯廷斯和皮科克(Evans, Hastings, & Peacock, 2000)用小写、斜体的英文字母 $a$,$b$ 和 $c$ 分别表示位置、尺度和形状参数。随机变量在基本参数的数目和种类上有所不同。例如,正态分布有两个参数:一个位置参数(均值)和一个尺度参数(标准差)。

## 子分布结构

分布在表示许多社会科学特别感兴趣的现象方面是很有用的。除了不平等之外,这些现象还包括总体的子群体结构。一般来讲,两种子群体结构是比较有兴趣的,而它们可能由分布的截删和删节【Censoring and Truncation】的运算版本来表示。在科茨、约翰逊和里德(Kotz, Johnson, & Read, 1982: 396)以及吉本斯(Gibbons, 1988: 355)的做法是把截删看成是用等级或百分数(或概率)点做的单位选择,而把删节看成是用变量值做的单位选择。因此,删节点是分割子分布的值 $x$;截删点则是分割子分布的百分数点 $\alpha$。例如,收入小于 20 000 美金的或大于 80 000 美金的子群体,每个都形成了一个删节子分布;总体顶部的 5% 和底部的 10%,每个都形成了一个截删子分布。

这两种子群体结构和布劳(Blau, 1974)的开创性观察之间有着特殊的联系。布劳观察到许多人类行为可上溯到其定量和定性性状的差异化运作(differential operation)。定量性状——基数性状(如财富)和序数性状(如智力或美丽)既可生产截删的子分布结构,也可产生删节的子分布结构。例如,子群体"富"和"贫"也许可参照收入量或人口的百分比产生。然而,定性性状,如族群、种族、语言或宗教可能与那些与子群体对应的不重叠的质性类别紧密相连,因而为产生子分布结构提供了基础。例如,在种姓制、奴隶制或隔离社会中,子分布结构感兴趣的可能是一种删节结构。在这种结构中,百分数与那些由一种质性的性状,如"奴隶"和"自由民"形成的子组有关。

## 理论工作中的分布

一种过程的理论分析或从一种理论进行预测推导常常会使用分布建模。这种方法使我们得以调查一个过程可能的变异性或与之相关的潜在分布,如收入分布形状的影响。这种方法的实施方案通常涉及选择恰当的建模分布和准备将它们用于特定的理论问题。

我们用三个理论问题来对这个问题作进一步的阐述。第一,基于由圣安塞尔姆提出的选择理论,调查本人收入的不平等效应;第二,基于一种由伯格和其他几位学者(Berger, Fisek, Norman, & Zelditch, 1977; Goode, 1978; Sørensen, 1979; Ridgeway, 2001)提倡的要素整合地位理论,调查在重要的定量性状上的等级地位的效应;第三,基于公正和比较过程理论,调查有价值的序数商品(ordinal goods)和基数商品(cardinal goods),与基数商品分布形状及它们在诸如幸福感和得到公正或不公正的报偿这样的结果上的效应(Jasso, 2001a)。

第二和第三种问题关注的是等级(总是在地位理论中,有时也在公正比较理论中)。等级,众所周知都直接由矩形分布表示(在涉及小群体问题中的离散矩形和涉及大社会问题中的连续矩形)。

第一和第三种问题主要包括基数商品的分布,而这些问题需要在有正面支持的分布中选择建模分布。一种方法是,考虑极端的行为,即是否存在最小或最大收入。这种方法会导出四种组合:(a)一个有非零最大下界和一个最小上界的变量;(b)一个有非零最大下界和无最小上界的变量;(c)一个有零最大下界和一个最小上界的变量;(d)一个有零最大下界和无上界的变量。我们用一个例子来作进一步的阐述。在好年景,一个原始社会的财富分布可以用第一种类型来建模;而在坏年景则可用第三种类型来建模。在发达社会,在战争即将结束时,在战胜国中,财富分布可用第二种类型来建模;在战败国中,则用第四种类型来建模。表1给出了这些性质的四种组合的每一种的例子。如表1所示,帕累托变量(Pareto variate),以及指数变量是非零最大下界和无最小上界的组合。对数正态是零最大下界和无最小上界的组合。另一种方法是,二次对称变量二者则有一个最大下界和一个最小上界。

**表1 物质商品的原型分布**

| 最大下界>0 | 最小上界 | |
|---|---|---|
| | 有 | 无 |
| 有 | 二次 | 指数<br>帕累托 |
| 无 | 幂函数 | 对数正态 |

注:在幂函数和二次变量中,最小上界也是最大值。对数正态和幂函数在零处有最大下界;$c>1$的幂函数也有一个最小的零。

在使用分布为第一和第三种问题的基量建模时,手中握有与用实质意义的参数、均值和不平等表示的变量的连带函数的公式是很有用处的。这样,第一步就是改变选择的变量,以使它们成为有两个参数的版本,而这两个参数则是算术均值和一般不平等参数。这两个参数都是不平等量度,都是单调函数。表2给出了表1列出的五个变量的累积分布函数、概率密度函数和分位函数的公式。

表 2　基数商品建模和基本连带函数

| 变量族 | 累积分布函数 | 概率密度函数 | 分位函数 |
| --- | --- | --- | --- |
| 指数 $x>\frac{\mu}{c},c>1$ | $1-\exp\left[-\frac{cx-\mu}{\mu(c-1)}\right]$ | $\frac{c}{\mu(c-1)}\exp\left[-\frac{cx-\mu}{\mu(c-1)}\right]$ | $\mu\left[\frac{1}{c}+\frac{c-1}{c}\left(\ln\frac{1}{1-\alpha}\right)\right]$ |
| 对数正态 $x>0,c>0$ | $F_N\left\{\frac{\left[\ln\left(\frac{x}{\mu}\right)+\frac{c^2}{2}\right]}{c}\right\}$ | $\frac{1}{xc\sqrt{2\pi}}\exp\left\{-\frac{\left(\frac{c^2}{2}+\ln\frac{x}{\mu}\right)^2}{2c^2}\right\}$ | $\mu\exp\left[cQ_N(\alpha)-\frac{c^2}{2}\right]$ |
| 帕累托 $x>\frac{\mu(c-1)}{c}$, $c>1$ | $1-\left[\frac{\mu(c-1)}{cx}\right]^c$ | $\left[\frac{\mu(c-1)}{c}\right]^c cx^{-c-1}$ | $\frac{\mu(c-1)}{c(1-\alpha)^{\frac{1}{c}}}$ |
| 幂函数 $0<x<\frac{\mu(c+1)}{c}$, $c>0$ | $\left[\frac{xc}{\mu(c+1)}\right]^c$ | $\left[\frac{c}{\mu(c+1)}\right]^c cx^{c-1}$ | $\frac{\mu(c+1)\alpha^{\frac{1}{c}}}{c}$ |
| 二次 $\frac{\mu(2+c)}{2}<x<\frac{\mu(2-c)}{2}$, $0<c<2$ | $\frac{1}{2}\left\{1-\cos\left[3\arccos\left(\frac{x-\mu}{\mu c}\right)-4\pi\right]\right\}$ | $\left(-\frac{6}{\mu^3c^3}\right)x^2+\left(-\frac{12}{\mu^2c^3}\right)x+\frac{3(c^2-4)}{2\mu c^3}$ | $\mu\left\{1+c\sin\left[\frac{\arcsin(2\alpha-1)}{3}\right]\right\}$ |

注：对所有的变量，$x>0$；其他的约束条件已在表中列出。$F_N(\cdot)$ 和 $Q_N(\cdot)$ 项分别表示单位正态变量累积分布函数和分位函数。不平等在帕累托和幂函数中是 $c$ 的递减函数，而在指数、对数正态和二次变量中则是递增函数。

在安塞尔姆式问题中，方案包括为每一模型的分布设立分位数函数，并求取本人收入对不平等度的一阶偏导数。结果包括预测本人收入随不平等的下降而上升的人在人口中的比例。该比例呈现很大的变化。例如，在收入是幂函数分布时，比例为37%；在收入分布是对数正态时为50%；而在收入是帕累托分布时，则为63%。在安塞尔姆式理论中，这些个人包括社会中最穷的那些人，因而对于增加了他们自己的收入也降低了收入的不平等的同样选择，他们是不能被冠以利己或利他的。

在地位和公正比较的问题中，一种早期的任务是分别寻找地位和正义比较结果的分布。这个任务使用地位和正义比较的公式（由各自的理论提供），给分布建模的公式应用变量变换、卷积和其他的技术（以标准分方式描述）来完成。在地位问题中，在一个好地位这一情形中，分布是很容易发现的，因为地位公式已经以等级的方式表示，它是正指数的二次函数（Sørensen，1979）；而在同时授予的地位的两种商品这一情形中，包括埃朗单位（Erlang）和至今尚未命名的族（Jasso，2001b：122）。已经求得的公正比较分布包括负指数、正指数、正态、Logistic 和对称与不对称的拉普拉斯（Jasso，2001a：683-685）。

地位问题和公正比较问题进一步分析

产生了各种各样可以检验的预测，例如迁出率、宗教机构的公共利益和处在战争状态的子群体之间冲突的严重性，都将随贵重的物质商品分布的不平等性的上升而上升。

当然，一个重大的挑战是求分布独立的结果。例如，使用来自概率分布研究的工具产生对称于零的公正比较分布的两个各自都为充分但非必要条件的结果。这两个条件是：(a)实际奖赏和公正奖赏是同一和独立分布的；(b)实际奖赏和公正奖赏是同一分布和完全负相关的(Jasso, 2001a: 683-685)。

## 经验工作中的分布

社会科学研究者经常关心他们的数据【Data】分布，特别是结果或因变量【Dependent Variable】的分布。尽管对于变量分布的研究往往出于各种各样的原因，但是其中最主要的两个原因是通过分布的分析得到知识和粗略地估计潜在的分布。

1.获得进一步的知识。分布参数的基本信息常常是不足够的，因而研究者会求助于绘图工具来帮助他们理解，以有可能对数据作进一步的操作。这些特性可以是在应用特定的编码【Coding】方案之前的直方图、茎叶图【Stem-and-leaf Display】或捕捉异常值【Outliers】和可能的强影响点【Influential Case】的个案的盒形图【Boxplot】，阐明分布不平等的洛伦兹曲线，或展示一个变量对于另一个变量的分布的散点图【Scatterplot】。现代统计学为我们提供了许多探索和审视分布的有用的工具(尤其在统计比较就是研究的目的时)。例如，相对分布法为研究者提供了一种比较两个收入分布的重要方法。

2.粗略估计潜在的分布。在社会科学实践中，常会将某种形式的回归【Regression】模型应用于我们的结果，这些模型的范围涵盖简单线性回归【Linear Regression】模型到非线性随机系数模型。借助这些做法，我们假定已经知道产生因变量的潜在机制。然而我们也经常会按惯例来指导我们的建模，而置分布的性质于不顾。在较好地捕捉潜在的分布时，即使是稍微作一下探索，也要花费很多的时间。图1给出了1991—2000年，英国家庭追踪调查(British Household Panel Survey)的妇女的样本分布。它用直方图代表数据的经验分布，而用线条表示三种理论分布。变量量度了抑郁，并记录了一个妇女报告“感觉比以前更坏了”的题项数。

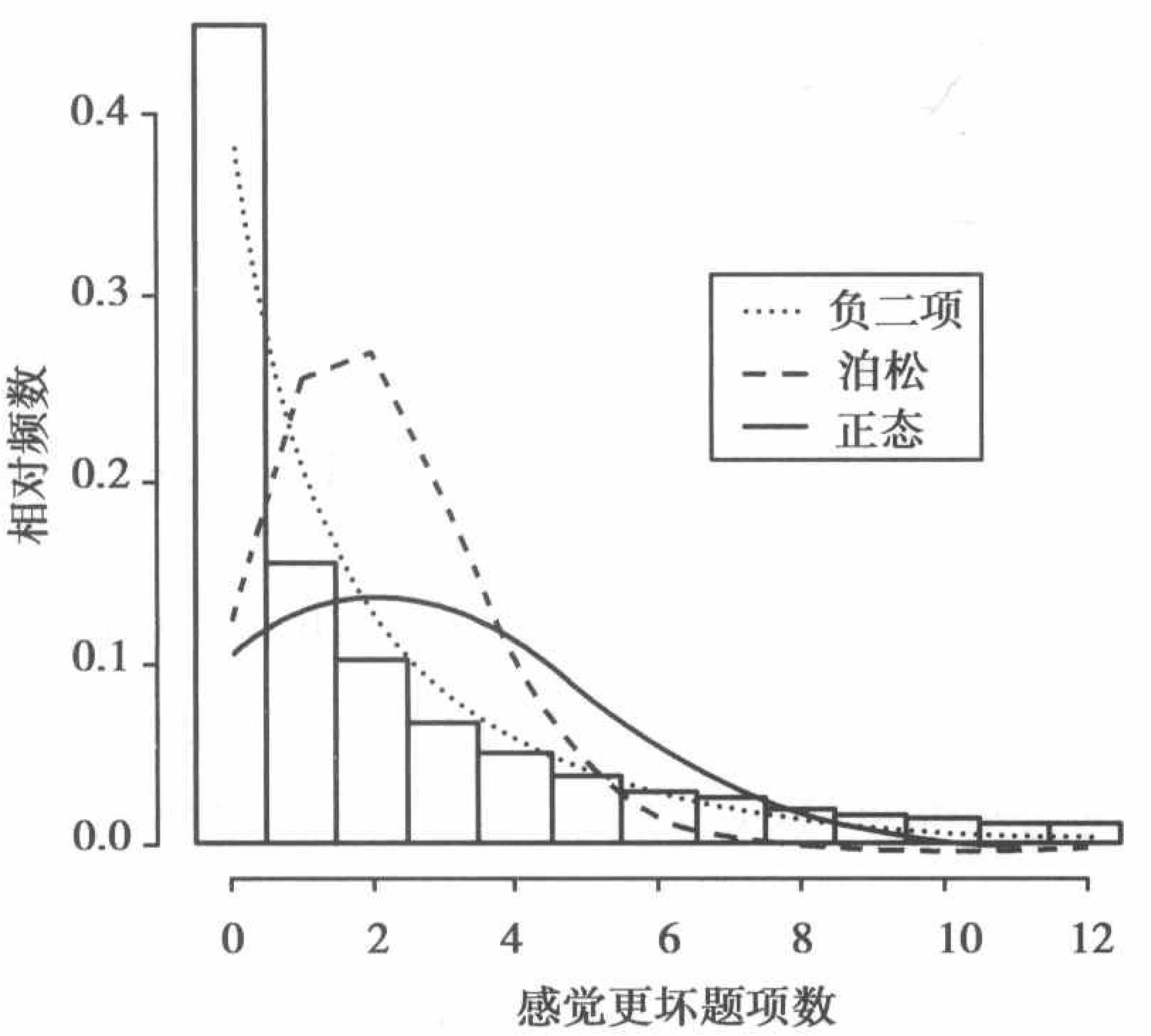



**图1** GHQ12 的分布，1991—2000 BHPS ($N$=7 855)

很清楚，正态分布并未能很好地表现数据。泊松分布可能更好一些，但是在这三种考虑的分布中，负二项分布由于能更好地表现数据，而优于其他两种，因此应该把它作为潜在的分布。

## 延伸阅读

有兴趣的读者，可阅读约翰逊和考兹所

撰的系列丛书中的《统计分布》卷和他们的修订版。在这些著作中,作者为我们提供了社会科学的理论和经验工作方面有关这一问题十分宝贵的资源。

——Guillermina Jasso
Tim Futing Liao
(沈崇麟译校)

## 参考文献


Berger, J., Fisek, H., Norman, R., & Zelditch, M. (1977). *Status characteristics and social interaction: An expectation states approach*. New York: Elsevier.

Blau, P. M. (1974). Parameters of social structure. *American Sociological Review*, *39*, 615-635.

Eubank, R. L. (1988). Quantiles. In S. Kotz, N. L. Johnson, & C. B. Read (Eds.), *Encyclopedia of statistical sciences* (Vol. 7, pp. 424-432). New York: Wiley.

Evans, M., Hastings, N., & Peacock, B. (2000). *Statistical distributions* (3rd ed.). New York: Wiley.

Gibbons, J. D. (1988). Truncated data. In S. Kotz, N. L. Johnson, & C. B. Read (Eds.). *Encyclopedia of statistical sciences* (Vol. 9, p. 355). New York: Wiley.

Goode, W. J. (1978). *The celebration of heroes: Prestige as a control system*. Berkeley: University of California Press.

Jasso, G. (2001a). Comparison theory. In J. H. Turner (Ed.), *Handbook of sociological theory* (pp. 669-698). New York: Kluwer Academic/Plenum.

Jasso, G. (2001b). Studying status: An integrated framework. *American Sociological Review*, *66*, 96-124.

Johnson, N. L., Kotz, S., & Balakrishnan, N. (1994). *Continuous univariate distributions*, *Vol. 1*. New York: Wiley.

Kotz, S., Johnson, N. L., & Read, C. B. (1982). Censoring. In S. Kotz, N. L. Johnson, & C. B. Read (Eds.), *Encyclopedia of statistical sciences* (Vol. 1, p. 396). New York: Wiley.

Ridgeway, C. L. (2001). Inequality, status, and the construction of status beliefs. In J. H. Turner (Ed.), *Handbook of sociological theory* (pp. 323-340). New York: Kluwer Academic/Plenum.

Sørensen, A. B. (1979). A model and a metric for the analysis of the intragenerational status attainment process. *American Journal of Sociology*, 85, 361-384.


# 分布无关统计量(Distribution-Free Statistics)

从最严格的意义上讲,分布无关统计量这一短语是指它的抽样分布【Sampling Distribution】不取决于一个随机样本从中抽取的那个总体【Population】。这就是说,无论其母总体的分布形式是什么样的,是正态、指数还是其他任何分布【Distribution】,统计量都有相同的分布。

最简单的分布无关统计量的例子是符号检验【Sign Test】统计量,它被定义为在一组 $n$ 对样本观察中的 $n$ 个差的加号数。不论成对的随机样本的总体分布形式是什么样的,这些差中的加号数都服从二项分布【Binomial Distribution】,有参数【Parameter】$p$,代表加号的概率【Probability】和 $n$(为样本量)。另一个分布无关的统计量,或者至少是近似分布无关的统计量是,从任何一个

方差【Variance】已知的连续总体抽取的,至少有 30 个观察的随机样本的标准化均值(或称 $z$ 分数【$z$-Score】)。根据中心极限定理【Central Limit Theorem】,这样一种标准化的样本均值近似正态分布。

实际上,任何一种分布无关统计量都允许研究者进行一种分布无关检验。这就是说,一个这样的检验,不必为了使检验被看成是有效的,而需要对样本抽取的那个总体的分布做任何假定。所谓有效检验是指它的第一类误差【Type I Error】的确切概率是规定的 $\alpha$ 值或其 $p$ 值【$p$ Value】是确切的。分布无关检验这一词已逐渐被非参数检验这一词所替代。非参数检验不仅包括了分布无关检验,而且还包括了任何虚无假设【Null Hypothesis】中不包括参数值的统计检验。这后一类检验包括拟合优度检验、趋势检验、自相关检验【Autocorrelation】和随机性检验等各种检验。非参数检验包括了一大批对小样本特别有用的统计分析方法。

——Jean D. Gibbons

(沈崇麟译校)

## 干扰项(Disturbance Term)

干扰项是统计模型中的误差【Error】项。如果干扰项是真正随机的【Stochastic】,那么它就是白噪声【White Noise】。

——Tim Futing Liao

(沈崇麟译校)

## 生活文献(Documents of Life)

生活文献是所有那些研究者可得到的,用于揭示人类社会及个人特征的文献。这种文献包括日记、信件、照片、生活故事、墓志铭、家居陈设、视频,以及如今的个人网站。在对这种方法较早的描述中,人文主义人类学家罗伯特·雷德菲尔德称这些文献是"个人文献",将其视为某种程度上文献作者对人类及其个人特点的表达,因此,通过这种方法,文献的读者能了解作者,以及他或她对文献所涉及的事情的看法(Gottschalk, Kluckhohn, & Angell, 1945, p.vii)。这些文献不是社会科学家对现实的二手建构,而是尝试用知情人自己的话来严肃地对待他们,是尝试进入知情人主观世界的一手描述。

——Ken Plummer

(杨可译 沈崇麟校)

*也可参见自传【Autobiography】、自传式民族志【Autoethnography】、日记【Diary】、生命历程研究【Life Course Research】。

### 参考文献

Allport, G. (1942). *The use of personal documents in psychological science*. New York: Social Science Research Council.

Gottschalk, L., Kluckhohn, C., & Angell, R. (1945). *The use of personal documents in history, anthropology, and sociology*. New York: Social Science Research Council.

Hoskins, J. (1998). *Biographical objects: How things*

*tell the stories of people's lives*. London: Routledge.
Plummer, K. (2001). *Documents of life-2: An invitation to a critical humanism*. London: Sage.

## 文献类型(Documents, Types of)

文献可以表现为任何一种在物理上体现的文本,或用手写或印刷在纸上的文本,例如一封书信或政府报告便是文献的原型。不过,文献的范围比这个要广泛得多。我们可以把文献分为书面的(文本是可见的语言文字)、音频的(文本是通过听觉渠道可以获取的声音)和视频的(文本是设计或图像表示的)等不同类型。文本可以用钢笔、铅笔、画笔、印刷机、录音机、计算机等来生成,也可以题写在陶土、石头、羊皮纸、纸张、磁盘或是电子显示屏上。文献的全部范围包括新闻、日记、邮票、名录、招贴、地图、照片、图画、唱片、录音带和计算机文件。

文献有时候被视为历史学家特有的原始材料,但文献在整个社会科学中的应用都很广泛。经典的社会科学家对文献资料的使用比调查法或观察法要多得多。马克思的《资本论》(*Capital*)大量引用了工厂巡视员的出版物,韦伯在《新教伦理与资本主义精神》(*The Protestant Ethic and the Spirit of Capitalism*)中使用了宗教传单和小册子,涂尔干在他的《自杀论》(*Suicide*)里也引用了许多官方统计数据。

我们可用原作者的身份和获取文献的条件,对文献进行最有效的分类。

作者身份可以根据文献是源于私人领域、公共领域还是公务领域再进行细分,而公务文献又可以按是源于政府还是私人机构再进行细分。这种区分很难精准,尤其是在前现代时期。比如说,在中世纪欧洲,因为世袭权威具有高度个人化的特点,因此在个人化的家庭文件、教会文件和政府文件之间做区分就没什么意义。

获取通道关注的问题是文献可供作者之外的人获取的条件。文献可以是“不公开的”,如果文献只是对非常有限的一群人开放,比如只对其作者和上级开放。如果文献是“受限的”,那么它就可供较广的人群使用,但必须要在文献作者或监管者规定的限制条件下使用。最自由的条件是“公开”,即通过档案或出版物获取文献的通道是公开的。这时会存在一条文档通道,其中的文档已作了分类整理,并可供更为广泛的公众使用:可能只需要很少的程序限制,例如需要注册或者获得一个读者证(参见数据档案【Data Archives】、定性数据存档【Archiving Qualitative Data】)就能获取文献。当文献出版以后,获取通道就几乎完全开放了,因为它们已完全向能负担得起的人开放了。

如表1所示,这两个维度可以结合起来产生12种不同种类的文献。不公开的个人文献(类型1)包括信件、日记、家庭账簿和其他家庭物品。这些文献种类通常只有拥有它们的个人或直接制造它们的家庭才能

**表1 文献类型**

| | | 作者身份 | | |
|---|---|---|---|---|
| | | | 正式 | |
| | | 个人 | 私人 | 政府 |
| 获取通道 | 不公开 | 1 | 5 | 9 |
| | 受限 | 2 | 6 | 10 |
| | 公开档案 | 3 | 7 | 11 |
| | 公开发表 | 4 | 8 | 12 |

获取。个人文献可以从一种获取类型变成另一种。有些个人文献如果被存放在公共档案馆,其开放获取的范围就变广了(类型3),有些可能得到出版(类型4)。例如,日记【Diary】本来是完全的不公开文献,但写日记常常是为了让更多读者见到。公共档案中保存了很多拥有土地的富裕家庭的记录,因此,这些文献也更易于获取了。不过,许多这类文献还是在私人手中——如果能一直保存下来——除非其持有者特别许可,否则公众就无法获取。

私人领域的正式文献是诸如企业、学校、医院和教堂等组织产生的文献。机密的机构文献(类型5)包括医疗档案、学历档案以及公司的人事档案,所有这些文献一般来说都只有相关人等,如负有行政或专业责任的人才能获取。受限的组织文献(类型6),例如,股东名册或房屋抵押贷款清单,通常只是在完全没有业务相关性时才能让研究者获取。对文献的查询和使用可能会受到特殊的限制。不过,对有的文献来说,法律要求必须保存在公共档案馆里(类型7):英格兰和威尔士的股东名册通常每年都要送交给公司登记处。出版的组织文献(类型8)也可能是按法律或其他官方要求而出版的,例如,公司要发表报告和财务报告,但是这类文献许多只是作为商业风险投资(commercial ventures)来出版的。这样的例子包括时间表和名录、参考书目、报刊、电影(参见研究中的影像【Film and Video in Research】),以及其他大众传媒产品。

政府文献,无论是地方的还是国家的,都有同样的获取类型,它们可能也是社会研究者可获得的文献中最多的一种。那些不公开的政府文献(类型9)包括犯罪记录和治安报告、地方当局的住房记录以及现今的税务记录。许多文献都受政府的保密法保护,除了直接的责任方之外,禁止未经授权就披露。有了这些法规,许多文献都是永远不公开的,尽管少数文件可能会向有限的使用者开放(类型10),也有些在公共档案中完全开放(类型11)。那些能在公共档案中查到的政府文件,通常都是在过了一段特定的时间、不再被视为保密文件之后才进入公共档案的。许多政府文件就是为了出版才生成的(类型12),包括议会的法案、皇家专门调查委员会的报告、统计报告(参见官方统计【Official Statistics】)和研究报告。

文献使用涉及的总原则和社会研究中其他领域中的原则差不多。尽管如此,文献资料的特别性质确实还是隐含着一些特殊的判断和技术。文献的有效使用取决于用四个标准对它们作出的评估,这四个标准即真实性、可信性、代表性和意义。

真实性标准要求必须对文献的可靠性和作者身份进行评估。可靠性评价意味着探查文献是原创的还是拷贝的,如果是拷贝的,那么究竟是对原作的拷贝,还是对拷贝的拷贝。拷贝,无论是手写、重新打字,还是拍照,都可能导致遗漏或不可读的文本,但即使是原件,也有可能会有印刷错误或缺失。一份"不可靠的"文献与理想化原创相去甚远,因为它在某些方面已经有了损伤。研究者在使用不可靠的文献之前,需要对之进行重构,重现可能的可靠文献原貌。文献受损越多,这个任务就越困难,在极端情况下,我们可能就无法重构原文献的可靠版本。

评价了可靠性之后就可以考虑作者身份了。比如说,了解据称属于某人的日记是否真的是由某人所写的就很重要。即使在很清晰明了的情况下,也可能会出现伪造作

者身份的问题。伪造文献可能是很重要的资料证据,但是必须要知道伪造的事实。作者身份可以通过内部和外部证据来评估。词汇和文风等内部证据,可以和纸张、墨水的化学检验,以及真实度评估等外部证据相结合,得出有关作者身份的判断。但是,在很多情况下,作者身份都不是那么明确的。在行政管理机构生产的正式文献中,任何署名的人,可能都只是起了很少一点作用。同样,书和报纸则是复杂的劳动分工的结果,署名作者的文字已经经过了文字编辑、副主编和主编的处理。

可信度问题关注的是文献的真实性和准确性。所有的文献都是选择性的,因为不可能构建出独立于特殊观点的叙述。尽管如此,叙述的可信度还是有高低之别,这取决于选择某种观点时的真实程度,以及是否按此观点进行了准确的报告。作者的真实性反映了他们的动机和他们进行选择的程度高低。人们为了自我合理化、为了宣传或欺骗、为了经济利益而写作,或者仅仅就是要尽量客观地写作,在各种情况下,评估文献受这些原因影响有多大都是很重要的。例如,正式文件可能是以"信息"的形式出现的,但其实际目的却是想说服特定位置上的人。同样,报纸是记者和其他拿报酬的人生产的,因而他们可能受制于出资方的政治压力。

即使真实地制造出来的文献也可能是不准确的。要评价一个报告的准确性,就要看它是在何种条件下编写的,尤其是要看作者和所报道的事件接触距离有多近。通常认为,原始资料——一手叙述——的准确性最好,因为这些文献给人的感觉是会把记忆流失、记录不全或无知所造成的准确性损失降到最低。但即使是一手观察也会遇到问题,那就是如何将观察用一种在日后能建构准确报告的形式记录下来。一般来讲,要这样做总是很困难的。比如说,要把谈话逐字逐句记录下来总是很难的。任何录音系统,尤其是机械系统都可能是一种打扰,因此可能会避免使用。观察者必须经常依赖记忆,因此,即使在原始资料中也会有回忆的准确性的问题。

评估代表性取决于对文献的保存和实用性的了解。了解所查询的文献是否代表了所有相关文献是很重要的,这取决于相关文献在多大比例上仍然得到了保存,以及是否能够为研究者所获取。如果文献要能存续,就必须以某种方式存储起来。存储可以指存在公共档案馆里,也可以指扔在卡片盒里。不是所有的文献都得被存储起来。许多公共和私人文献在生成之后不久就被销毁了,还有一些文件保存一段时间之后也会被销毁。计算机越来越多地被用于建立和保存文献,这意味着在很多情况下不会生成历史文献:文献都是实时编辑和更新的,因此永远只有最后的版本可用。对于以备份形式归档以备将来研究之需的计算机文件,很难知道其内容与原始的连续数据库有多大程度的关联。许多个人文献都没有保留下来,而保存下来的个人文献之所以能保存,是因为其持有者认为它在某些方面有不同寻常之处。在现代科层制机构所生产的大量文献中,只有很小一部分能保留下来。正式文献是在使用中被存储起来的,可能在被移到档案库之前就被删掉了。随着时间的流逝,由于文献解体或者周期性的"清理",能保存下来的文献数量可能会减少。这种损失有时会随机发生,但也可能会涉及某些未知的系统过程。

保存下来的正式文件是否可得,还得考

虑其机密性和官方的保密性。政府文件通常在经过特定的一段时间之后再进入公共领域，但有的文献可能永远受限。个人和私人领域的问题可能更大，尤其是对家庭或家居文献而言，几乎没有什么法律对其使用权作出规定。个人文献也常常缺少索引和日历，而通常官方文献就是借助于索引和日历来评估其代表性的。除非是保存在家庭档案中，除了非常富裕的家庭之外，这种情况都很罕见，私人家居物品既没有档案，也没有目录。

尽管文献必须评估其真实性、可信性和代表性，但它们不必是完全真实、可信或有代表性的。重要的是，研究者知道他或她所处理的文献在一定程度上是不真实的、不可信的或者不具有代表性的，这样接下来才可以考虑其意义的问题。

文献的意义问题源于两个层面：文字层面和阐释层面。对文献的文字理解牵涉文献的物理可读性，文献使用的是否是可读的语言，以及诸如日期这样的问题。阅读褪色或腐烂的文献可能需要显微镜、红外光等技术，阅读计算机文件总是需要特定的硬件和软件。阅读音频和视频资料按常规都需要特定的技术。从原则上讲，文字意义的问题是可以从实际出发加以解决的，给研究者留下一个有意义的文本进行分析。

正是在这里出现了更为基本的阐释意义的问题，而且事涉复杂的理论问题。阐释是一个诠释学的任务，涉及对社会和文化环境，以及构建文本的话语形式的理解。我们必须抓住其潜在的观点，因为这是文本中的个人概念赖以获得意义的基础。对真实性、可信性和代表性的评估揭示了文献产生的环境，而文献则必须在这种环境中予以解释。

本词条中的文本分析方法远不够详尽，不过两个主要的方法则是定量的内容分析【Content Analysis】和定性的语义方法（也请参见定性内容分析【Qualitative Content Analysis】）。前者包括了技术和统计分析等方法，后者包括解码和解构等方法。不过，基本上所有的阐释最终都是一种定性的工作，不仅如此，即使是定量的内容分析也要依赖于通过一种经常是暗示的记号语言过程（often-implicit semiotic process）来进行的意义评估。

评估文献时，真实性、可信性、代表性和意义这四个标准不能被视为分离和独特的阶段。它们是相互倚赖的，研究者不可能在不考虑其他几个标准的情况下充分考虑其中任何一个标准。从这个意义上讲，对文献的评价是个永远也不会结束的过程。研究者提出的阐释意义是一种尝试性的、临时性的判断，必须不断地加以修正，当新发现和新问题出现时，就不得不对文献进行重新评估。理解这一点，对研究者明智地使用文献至关重要。

——John Scott

（杨可译　沈崇麟校）

## 参考文献

Scott, J. (1990). *A matter of record: Documentary sources in social research*. Cambridge, UK: Polity.

# 双盲程序(Double-Blind Procedure)

双盲实验程序是设计用来消除由研究参与者、实验者和数据分析者的期望造成的偏倚。调查研究者在自己的实验设计中普遍采用双盲的三个层次中的一个。在单盲实验中,参与者不清楚自己被分配在了处理组还是控制组。在双盲实验中,参与者和实验者都不清楚组的具体分配。最后,在三盲实验中,参与者、实验者和那些负责数据输入和分析的人都不知道每次研究试验的处理和控制条件。

双盲程序是必要的,因为参与者不仅会对实验的刺激作出反应,而且也会对实验假设【Hypothesis】的印象作出反应。如果实验的操作处理是一览无遗地,那么参与者就可能朝他们认为会得到赞许的反应方向给出自己的反应。在这种情况中,实验量度的答案代表预期的治疗效应和实验设置的需求特性的一个未知的组合。

更为微妙的偏倚【Bias】源潜藏在实验者和参与者之间的互动中。在实验者期望在某一种条件下来自参与者的某种反应时,实验者可能会通过(常常并非有意而为地)用非语言的提示、额外的鼓励或提起注意等手段来影响参与者,人为地产生或加强这种反应。从这些行为导致的偏倚称为实验期望效应。因为参与者和实验者偏倚与实验的操作系统相关,因此它们可能会对实验结果的解释产生混淆。这同一种类型的期望效应可能使人们在处理数据时,自觉或不自觉地偏向于自己期望的方向。

双盲程序也称为“戴面具的条件”或“实验上纯真(experimentally naive)”的程序,经常在医学研究中使用。例如,在药物试验中,将一种外表和实验的药片一样的糖片用于控制组。实验的药片和糖片处理由一个与参与者没有打过交道的调查研究者随机分配。药片的同一性使参与者和给予药片的实验者都不知道每次试验的性质。双盲程序还有一个优点,那就是可以防止那些做实验的人,因同情贫困的病人而给他们实验的药物,而不是安慰剂,从而扭曲试验的结果。

许多社会科学的调查都涉及处理问题,与药片不同的是,这些处理不能被设计成看起来相同的处理,也不能对主体加以控制。我们来看一个比较电话和电子邮件信息对态度变化的效应的研究。在这里,参与者和实验者都完全清楚研究涉及的是一个电话还是一个电子邮件。在这种情形中,双盲则通过使参与者和实验者都不知道假设,或尽可能地减少参与者和知道假设的实验者之间的联系来近似。例如,在因变量【Dependent Variable】的数据收集之前,由第三方来实施处理。总之,一个“失明”的协议【Protocol】有助于调查研究者最大限度地减少系统的人为偏倚。这样的偏倚都是在参与者、实验者和数据分析者清楚实验处理的方法和目的时引起的。双盲程序在描述处理和结果之间的因果关系时,给调查研究者以更多的信心。

——Donald P. Green
Elizabeth Levy Paluck
(沈崇麟译校)

## 参考文献

Aronson, E., Carlsmith, M., & Ellsworth, P. C. (1990). *Methods of research in social psychology*. New York: McGraw-Hill.

Kazdin, A. E. (2003). *Research design in clinical psychology* (4th ed.). Boston: Allyn & Bacon.

# 双重度量(Dual Scaling)

这是一种双向表(例如列联表【Contingency Table】,多选择答案的考题表)中的分类数据的多维度量化方法。"双重"这一个词指它是一张表的行和列的对称尺度。双重度量为多选项题项的选项确定权重。它以由选择的选项的加权之和给出的这个分数使辨别力达到最大,从而使内部的一致性也达到最大这样一种方式来确定应试者的分数。然后再将这些选项的权数表达为选择这些选项的考生的加权分数之和。这一性质常被称作路易斯·哥特曼内部一致性原则。弗雷德里克·洛德证实,诸如这样的加权方式,最大化了导出分数内部一致性的信度。

## 历史发展

早期分类数据的多维度量化工作,可追溯到马里奥·理查森和弗雷德里克·库恩在1933年使用的方法,保罗·霍斯特在1935年把它称为互反平均数(reciprocal averages)法。赫希菲尔德在1935年提出了一种数学上等价的方法,詹姆斯·林格斯在1964年把它称为联立线性回归。罗纳德·费希尔在1940年提出了分类数据的判别分析【Discriminant Analysis】,而貌(Maung)则在1941年证明费希尔的量化法可以由四个不同的目标函数以公式来等价表示。而貌的论文则适用于列联表,哥特曼在1941年提出了三个在数学上等价的、多项选择数据的量化方法,并在1946年将它扩展到成对比较和等级次序数据。1946年,分类数据的量化方法便有了坚实的基础。1950年,林知己夫在日本启动了一个叫作Hayashi量化理论的项目。该理论的三型便相当于列联表的双重度量。1960年代早期,让-保罗·贝内泽在法国开始了一个研究,后来发展成为列联表的对应分析【Correspondence Analysis】和多项选择数据的多重对应分析。林知己夫和让-保罗·贝内泽两人有着许多追随者,他们被称为数据分析的林知己夫学派和让-保罗·贝内泽学派。类似的共同努力在1960年代后期形成了荷兰的莱登研究组(以让·德莱乌为首)和加拿大的多伦多研究组(以希祖海科·尼什撒多为首)。

一直以来,这种方法曾经有过许多名字,如Guttman度量(Guttman)、互反平均数法(Richardson, Kuder, Horst)、联立线性回归(Hirschfeld)、Fisher适当得分(Fisher)、Hayashi量化理论三型(Hayashi)、最优度量(R. Darrell Bock)、对应/多重对应分析(Benzécri, Brigitte Escofier)、双标图(Ruben Gabriel, John Gower, David Hand)、同质性分析(de Leeuw, Willem Heiser, Jacqueline Meulman)和双重度量(Nishisato)等。双重度量这一普及的名字是在1976年,为了回应流行的最优度量不足以专门用来描述这种方法的批评而提出的。根据缪尔曼(Meulman, 1998)提出的观点,双重度量是一种"用于分类数据的多维度分析的全面的框架"(p. 291),因为它处理的分类数据的范围比诸如对应分析和最优分析这些主要应用于发生率数据的方法更广。

## 基本公式

这一方法为分类数据采用奇异值分解。我们来看一张双向列联表,该表有在行 $i$ 和列 $j$ 的频数 $f_{ij}$。这一元素的奇异值分解可表达为:

$$f_{ij} = \frac{f_{i.} f_{.j}}{f_{..}}[1 + \rho_1 y_{i1} x_{j1} + \rho_2 y_{i2} x_{j2} + \cdots + \rho_k y_{ik} x_{jk} + \rho_K y_{iK} x_{jK}] \quad (1)$$

式中的 $f_{i.}$ 是行 $i$ 的边缘频数;$f_{.j}$ 是列 $j$ 的边缘分布;$\rho_k$ 是第 $k$ 个奇异值($\rho_1 \geq \rho_2 \geq \cdots \geq \rho_k$);$y_{ik}$ 是行 $i$ 的第 $k$ 个解的最优权数;而 $x_{jk}$ 则是列 $j$ 的第 $k$ 个解的最优权数。对应于括号内 1 的解称为平凡解,一般都会在分析中删除,而它正是在行和列在统计上独立时的期望频数。其他解都称为正常解。对于任何正常解,比如说解 $k$,存在着双重关系(Nishisato, 1980)或转换公式(Benzécri et al, 1973):对于一个 $m \times n$ 的数据矩阵

$$y_{ik} = \frac{1}{\rho_k} \frac{\sum_{j=1}^{n} f_{ij} x_{jk}}{f_{i.}} \text{ 和 } x_{jk} = \frac{1}{\rho_k} \frac{\sum_{i=1}^{m} f_{ij} y_{ik}}{f_{.j}} \quad (2)$$

方程(2)无非是行在列上或列在行上的联立线性回归而已。它也指出了互反平均数法的偏倚:加权列均值的计算始于任意的行权数,接着它便作为计算加权行均值的权数。这样的互反平均路数总是收敛(Nishisato, 1980)于行权数和列权数的最优集合,其比例常数就是奇异值。尼什撒多(Nishisato, 1980)把权数 $y$ 和 $x$ 称为赋范权数,而把 $\rho_y$ 和 $\rho_x$ 称为投影权数,它们也被分别称为标准坐标和主坐标(Greenacre, 1984)。在几何学上,$y$ 和 $\rho_x$ 跨越同一空间,而 $\rho_y$ 和 $x$ 也跨越同一空间。行空间和列空间之间的差异角度则由奇异值的余弦给出(Nishisato & Clavel, 2003)。

**表 1**

| 聚会 | 计划 | 等级 | | | | | | 优势矩阵 | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | (1) | (2) | (3) | (4) | (5) | (6) | (1) | (2) | (3) | (4) | (5) | (6) |
| 应试者 | 1 | 6 | 1 | 5 | 4 | 3 | 2 | −5 | 5 | −3 | −1 | 1 | 3 |
| | 2 | 6 | 1 | 5 | 2 | 4 | 3 | −5 | 5 | −3 | 3 | −1 | 1 |
| | 3 | 3 | 5 | 2 | 4 | 1 | 6 | 1 | −3 | 3 | −1 | 5 | −5 |
| | 4 | 3 | 4 | 2 | 6 | 1 | 5 | 1 | −1 | 3 | −5 | 5 | −3 |
| | 5 | 5 | 3 | 1 | 4 | 6 | 2 | −3 | 1 | 5 | −1 | −5 | 3 |
| | 6 | 2 | 6 | 3 | 5 | 4 | 1 | 3 | −5 | 1 | −3 | −1 | 5 |
| | 7 | 1 | 2 | 4 | 5 | 3 | 6 | 5 | 3 | −1 | −3 | 1 | −5 |
| | 8 | 4 | 3 | 2 | 6 | 5 | 1 | −1 | 1 | 3 | −5 | −3 | 5 |
| | 9 | 2 | 1 | 4 | 5 | 3 | 6 | 3 | 5 | −1 | −3 | 1 | −5 |
| | 10 | 6 | 1 | 4 | 3 | 5 | 2 | −5 | 5 | −1 | 1 | −3 | 3 |

列联表和多选择数据被表示为回答类型，它们都是发生率数据的例子，它们被分解成方程(1)的表达式。相反，优势数据包括等级次序数据和成对比较数据，它们的观察都先按下述方式进行编码：

$$ {}_if_{jk} = \begin{cases} 1 & \text{如果应试者 } i \text{ 选择目标 } j \text{ 到 } k \\ 0 & \text{如果应试者 } i \text{ 作出同分的判断} \\ -1 & \text{如果应试者 } i \text{ 选择目标 } k \text{ 到 } j \end{cases} \quad (3) $$

这些编码后的观察，通过下式

$$ e_{ij} = \sum_{j=1}^{n} \sum_{k=1(j \neq k)}^{n} {}_if_{jk} \quad (4) $$

转换成优势数。

优势数矩阵有行元素之和为零这样一种性质。因此，为了达到应试者的权数和目标权数，例如，在使用互反平均法时，我们都要明白优势矩阵的每一个格都是基于 $n-1$ 个回答的。

## 发生率数据

1933 年，尼什撒多将分类数据分为了两种类型，即发生率数据和优势数据。发生率数据由 1,0 或频数主组成，存在一个平凡解，使用卡方度量，而且即使在所有的变量都完全相关时，为了完整地描述数据，一般也需要有一个以上的维度。像发生率数据这样一种类型的数据，上面提到的用于它的量化方法都已经开发出来了。诸如列联表、多项选择数据和排序数据这样的发生率数据的分析实例，请参见吉非（Gifi，1990）、格里纳克（Greenacre，1984）和尼什撒多（Nishisato，1980，1994）的实例。

## 优势数据

优势数据由大于、等于或小于这样的回答组成，不涉及平凡解，使用欧几里得度量，在变量完全相关时，用一个维度便足以描述数据。因为双重度量也可处理优势数据，所以让我们来看一个例子。我们对十位应试者根据他们对六种圣诞聚会计划的偏好做了排位。这六种偏好分别是：(1)在白天团体聚餐；(2)下班之后在酒吧-饭店玩耍；(3)在地区饭店享用价格适当的午餐；(4)在旅馆举行晚宴；(5)下班后在某人家中聚餐；(6)在一家高档的饭店进豪华午餐。数据来自《心理测量学》（*Psychometrika*，Nishisato，1996），而相应的优势数据则在表 1 中。五个正常解的相对方差分别为0.45，0.27，0.14，0.08 和 0.06。我们来看一下前两个正常解和应试者的赋范权数和聚会计划的映射权数 $\rho x$ 的联合图，练习导致一个克莱德·库姆斯的多元展开问题的解（Nishisato，1994）（参见图 1）。从这张图中，我们可计算应试者和聚会计划之间的距离。排列每个应试者内的距离，得到近似于输入排位的排位-2（参见表 2）。我们来看一下应试者 2，距离排位恰好复制了该应试者的计划排位。

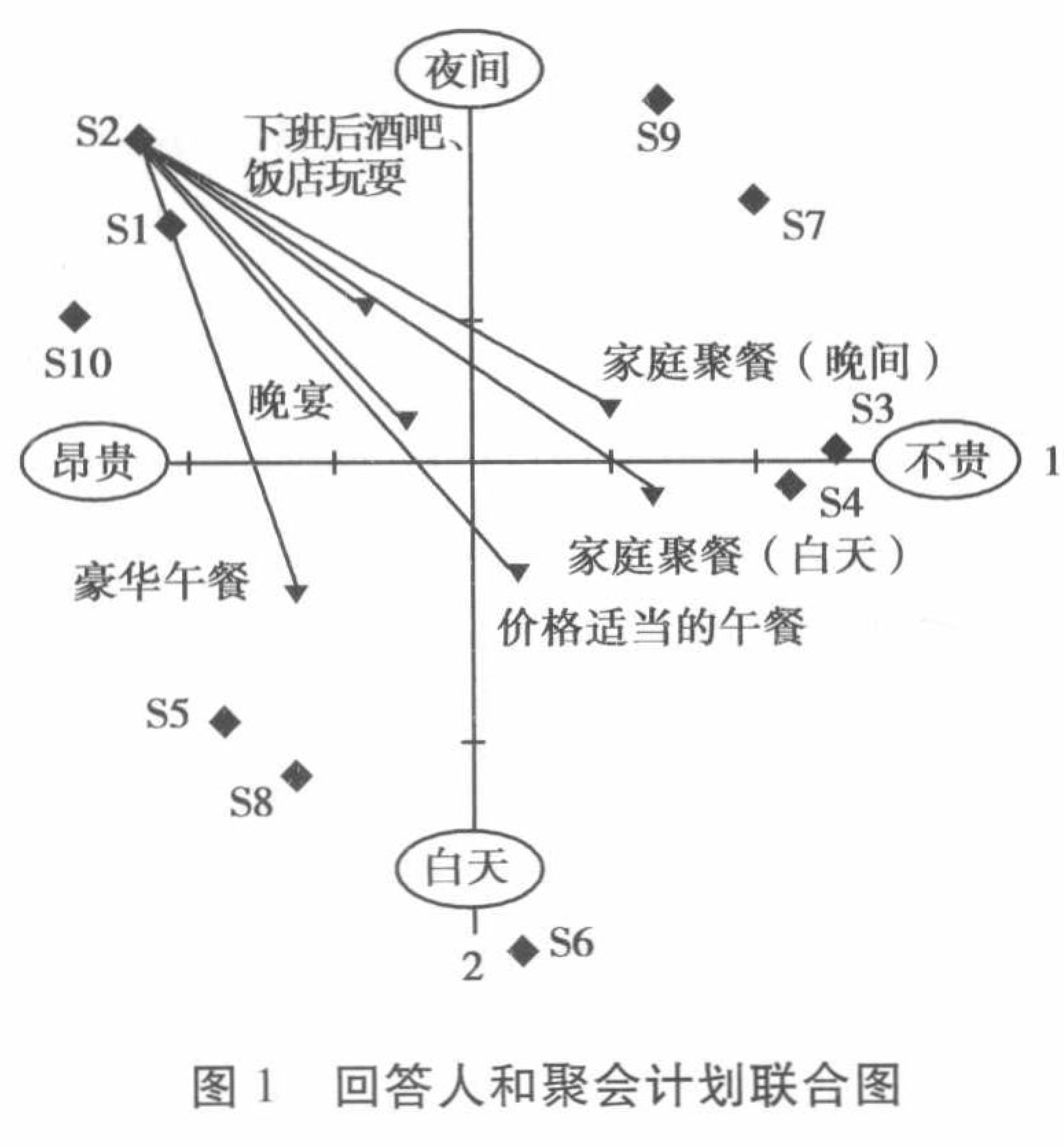



图 1　回答人和聚会计划联合图

表 2

| | 距离平方 | | | | | | 排位-2 近似 | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (1) | (2) | (3) | (4) | (5) | (6) |
| S1 | 1.94 | 0.73 | 1.72 | 1.06 | 1.67 | 1.37 | 6 | 1 | 5 | 2 | 4 | 3 |
| S2 | 2.20 | 0.98 | 2.02 | 1.36 | 1.91 | 1.71 | 6 | 1 | 5 | 2 | 4 | 3 |
| S3 | 0.67 | 1.75 | 1.21 | 1.53 | 0.81 | 1.98 | 1 | 5 | 3 | 4 | 2 | 6 |
| S4 | 0.49 | 1.63 | 1.01 | 1.38 | 0.69 | 1.79 | 1 | 5 | 3 | 4 | 2 | 6 |
| S5 | 1.72 | 1.58 | 1.18 | 1.27 | 1.77 | 0.54 | 5 | 4 | 2 | 3 | 6 | 1 |
| S6 | 1.70 | 2.38 | 1.37 | 1.95 | 1.96 | 1.51 | 3 | 6 | 1 | 4 | 5 | 2 |
| S7 | 1.10 | 1.42 | 1.55 | 1.45 | 0.89 | 2.14 | 2 | 3 | 5 | 4 | 1 | 6 |
| S8 | 1.61 | 1.70 | 1.07 | 1.33 | 1.71 | 0.76 | 4 | 5 | 2 | 3 | 6 | 1 |
| S9 | 1.40 | 1.28 | 1.74 | 1.45 | 1.11 | 2.18 | 3 | 2 | 5 | 4 | 1 | 6 |
| S10 | 2.13 | 1.02 | 1.79 | 1.21 | 1.91 | 1.25 | 6 | 1 | 4 | 2 | 5 | 3 |

其他的例子要求两个以上的维度。总的来讲，排位-2 近似是接近排位-5 数据的。统计图 1 告诉我们，排位中的个体差异，受聚会费用和聚会时间的影响。请读者观察它们对特定计划的接近程度。

——Shizuhiko Nishisato

（沈崇麟译校）

## 参考文献


Benzécri, J. P., et al. (1973). *L'analyse des données: II. L' analyse des correspondances*. Paris: Dunod.

Gifi, A. (1990). *Nonlinear multivariate analysis*. New York: Wiley.

Greenacre, M. J. (1984). *Theory and method of correspondence analysis*. London: Academic Press.

Lebart, L., Morineau, A., & Warwick, K. M. (1984). *Multivariate descriptive statistical analysis*. New York: Wiley.

Meulman, J. J. (1998). Book review of W. J. Krzanowski, & F. H. C. Marriott, Multivariate analysis. Part I. Distributions, ordinations, and inference, *Journal of Classification*, *15*, 287-293.

Nishisato, S. (1980). *Analysis of categorical data: Dual scaling and its applications*. Toronto: University of Toronto Press.

Nishisato, S. (1994). *Elements of dual scaling: An introduction to practical data analysis*. Hillsdale, NJ: Lawrence Erlbaum.

Nishisato, S. (1996). Gleaning in the field of dual scaling. *Psychometrika*, *61*, 559-599.

Nishisato, S., & Clavel, J. G. (2003). A note on between-set distances in dual scaling and correspondence analysis. *Behaviormetrika*, *30*, 87-98.

## 虚拟变量(Dummy Variable)

一套虚拟变量让我们得以把分类的解释变量设定为统计分析中的自变量。有些我们想加入统计模型的信息是以分类表来组织的,代表了一套名义类别(nominal categories)。婚姻状态、性别、种族/民族、政党或者工业部门都可以用一套完备而互斥的类别来表示,每一个观察值都可以分配给一个单独的类。在最简单的情况下,我们有两个类(例如男性和女性),我们可以用一个单独的虚拟变量来代表它。如果是男性,其编码为"1",如果不是男性,编码则为"0"。这种类型的编码【Coding】被称为二分编码,表明在某个具体的类别中的某种属性或成员资格的出现。对多于两个类别的分类来说,如果我们从策略上给 $j-1$ 个变量编码,就可以充分地代表分类信息,并生成提供可阐释信息的系数(Hardy,1993;Hardy & Reynolds,2004)。

在设定方程右侧的变量(自变量)时,研究者可能想要了解,比如说,婚姻状态的差异和某个因变量(如抑郁症状)之间的相关程度如何。尚待解决的问题是,如何能不做不真实的分布假设来表示有 $j$ 个类别的分类表包含的信息。例如,在一个有关美国成年男女的样本中报告了婚姻状态,我们可以区分出诸如:(1)已婚;(2)离婚;(3)分居;(4)丧偶;(5)从未结婚。如果我们只是简单地在某个统计模型中包括一个"婚姻状态"的变量,而不以任何方式对分配给不同婚姻状况的 1~5 的编码进行某种修改的话,那么我便是假定婚姻状态是定距测量的,也就是说,在"已婚"和"离婚"之间 1 个单位的数值差等于"丧偶"和"分居"之间 1 个单位的数值差,但只等于"丧偶"和"离婚"之间数值差的一半,而且这些类别的排序还都很恰当,在该取值范围中,"已婚"的取值最低,"从未结婚"的取值最高。这种方法显然是荒谬的;每个类别的取值和类别的排列顺序都是任意的。我们需要一种符合定类分布特点的方法,在这个案例中,也就是说,我们需要一种把独特的类别分离开来的数值方式,这样我们就可以用计数(即频次)和样本划分来提供合适的估计值。

尽管编码方法有很多,但最常用的方法有三种:二分编码【Binary Coding】、对照编码【Contrast Coding】和效果编码【Effects Coding】。回到婚姻状态这个例子上来,我们需要用四个虚拟变量来充分代表五类别分类中的信息。采用二分编码,并选择"已婚"作为我们的参照组(剩下的类别将与之进行比较的组),可以把四个虚拟变量命名为"离婚""分居""丧偶"和"从未"。婚姻状态符合变量名的被调查者就在这个虚拟变量上编码为"1"。比如说,现在处于离婚状态的被调查者就在"离婚"上编码为"1",在剩下的三个变量上编码为"0";从未结婚的被调查者就在"从未"这个变量上编码为"1",剩下的编码为"0",以此类推。这样,四个虚拟变量就完全捕捉到了五类别定类变量含有的信息这一点,则为在所有四个变量上编码都为"0"的任何调查者必然是已婚的这一点所证明。

无论采用何种编码方案,确定分类变量——这个例子中是婚姻状态——是否提高了模型拟合度的适当的推论检验,都是比较有没有虚拟变量的模型拟合的 $f$ 检验。虚拟变量拓展了广义线性模型【General Linear Model】的灵活性,让我们可以在更广的范围内检验假设,包括直接效应和间接效应、交互【Interaction】以及不同的函数形式。

在统计模型中,二分编码的虚拟变量也

可以被设定为因变量。在一套用于有限应变量模型估计的方法中,包括了二元概率单位(probit)和二分几率比(logit)模型(Long & Cheng,2004)。离散时间的事件史【Event History】模型也使用二分编码的因变量来说明某个事件会不会发生(Allison,2004)。二分编码的因变量也在很多选择模型中使用,例如 Heckman 二阶段程序和内生转换模型(Fu,Mare, & Winship,2004)。

——Melissa A. Hardy

(杨可译　沈崇麟校)

## 参考文献


Allison, P. D.(2004). Event history analysis. In M. A. Hardy & A. Bryman(Eds. ), *Handbook of data analysis*. London: Sage.

Fu, V., Mare, R., & Winship, C.(2004). Sample selection bias models. In M. A. Hardy & A. Bryman (Eds. ), *Handbook of data analysis*. London: Sage.

Hardy, M. A. (1993). *Regression with dummy variables*. Newbury Park, CA: Sage.

Hardy, M. A., & Reynolds, J.(2004). Incorporating categorical information into regression models: the utility of dummy variables. In M. A. Hardy & A. Bryman (Eds.), *Handbook of data analysis*. London: Sage.

Long, J. S., & Cheng, S.(2004). Regression models for categorical outcomes. In M. A. Hardy & A. Bryman (Eds.), *Handbook of data analysis*. London: Sage.


# 持续期分析(Duration Analysis)

参见事件史分析【Event History Analysis】。

# Durbin-Watson 统计量(Durbin-Watson Statistic)

Durbin-Watson 统计量($DW$ 或 $d$)是一种普遍使用和例行报告的时间序列回归(time-series regression)模型误差中存在的一阶自或序列相关【Serial Correlation】的诊断检验的统计量。该统计量用我们想求其 $\hat{\mu}$ 的回归的残差【Residuals】计算,其计算公式为:

$$DW = \frac{\sum_{t=2}^{n}(\hat{\mu}_t - \hat{\mu}_{t-1})^2}{\sum_{t=1}^{n}\hat{\mu}_t^2}$$

这一检验的虚无假设【Null Hypothesis】是在回归误差【Errors】中没有一阶自相关,它有两种备择的方法:正的和负的一阶自相关。

它可以直接用一阶自相关 $\rho$ 表示的检验统计量来阐释:

$$DW \approx 2(1-\rho)$$

以这种方式来书写的统计量,使我们很容易看清检验统计量的值和模型残差中的自相关程度之间的关系。特别是,如果自相关不存在($\rho=0$),那么 $DW$ 就近似 2;如果存在完全的正的自相关($\rho=1$),那么 $DW$ 就近似 4;而如果存在完全的负的自相关($\rho=-1$),那么 $DW$ 便近似 0。

检验的临界值【Critical Values】随一个

样本【Sample】量的函数变化，而它的自由度【Degrees of Freedom】则与感兴趣的回归有关联。一般 *DW* 检验都按正的一阶自相关的备择假设计算，因此我们要寻找的是显著小于 2 的 *DW* 值。假如难以得到虚无假设下的 *DW* 统计量的分布，那么每一个备择假设都需要有两个临界值。

只有在我们关心的回归右侧不存在滞后【Lagged】的因变量时，这一检验才是合适的检验。如果回归含有滞后的因变量，那么可资选择的检验包括德宾的备择检验、德宾的 *H*（Durbin's *H*）或拉格朗日乘子检验（Lagrange multiplier tests）。同样，*DW* 将无法探测高阶的自相关。为了检验误差中的二阶或更高阶的自相关，我们可以使用拉格朗日乘子检验。

*DW* 统计值显著则表明存在自相关，它将使我们关心的回归参数【Parameters】假设【Hypothesis】检验失效。一个显著的 *DW* 统计值一般可作为模型【Model】误设的证据，但也可能表明的是误差本身的序列相关（Wooldridge，2000）。适当的补救措施应根据问题的成因而定。模型也许要求在方程的右侧包含一个滞后的因变量或其他动态的回归量。有时，我们可能需要通过变换的变量，用可行的广义最小二乘【Generalized Least Squares，GLS】估计回归重新估计原回归，以去除自相关，并进而估计得到的回归。

我们用一个例子来作进一步的阐释。假设我们希望给从 1921 年到现在的美国妇女投票率（$v_t$），作为妇女的平均教育程度（$e_t$）和妇女在外就业的百分比（$\omega_t$）的函数建模。为了检验一阶序列相关，我们先要求下面这一模型的残差：

$$v_t = \beta_0 + \beta_1 e_t + \beta_2 \omega_t + \mu_t$$

然后，再将这些残差用于计算像上面这样的 *DW* 统计量。对于一个样本量为 $t = 45$ 和三个自由度的（两个回归量加一个常数）回归，正的自相关的备择法的 5%的临界值是 1.45（低端）和 1.62（高端）。假定估计的 $DW = 1.8$，那么我们将不能拒绝不存在正的自相关的虚无假设，并进而推断可能不存在一阶自相关，如果估计的 $DW = 1.4$，我们将拒绝没有正的自相关的虚无假设。对于所有在 1.45～1.62 的值，检验是不确定的，因而我们将推断可能有一阶自相关存在。

——Suzanna De Boef

（沈崇麟译校）

## 参考文献


Durbin，J.，& Watson，G. S.（1950）. Testing for serial correlation in least squares regressions I. *Biometrika*，*37*，409-428.

Wooldridge，J. M.（2000）. *Introductory econometrics：A modern approach.* Cincinnati，OH：South-Western Thomson Learning.


# 偶体分析（Dyadic Analysis）

在社会研究中常常要测量成对的人（如已婚夫妻、室友、朋友和同事）。成对数据的统计分析称为偶体分析，偶体数据分析比单体数据分析要复杂得多（Kashy & Kenny，2000）。我们需要对数据的互依（interdependence）和分析单位【The Unit of Analysis】予以特别

的考量。

一般情况下，两个人的分数相关便会产生一种称为相依的现象。分数的相关使那些用于以个人作为分析单位的显著性检验【Significance Testing】失效。来自检验的 $p$ 值可能大于或小于它们似乎是的数值。这种分数相依甚至有可能是负的，这样，偶体中的两个成员相互之间的差别便会大大大于两个随机选择的人。在成员可与某个变量区别开来时，相依程度可用皮尔逊相关系数【Pearson Correlation Coefficient】来测量。例如，异性的已婚夫妇，可以用他们的性别来区分。在成员无法区分的（如同事）时候，应用组内相关【Intraclass Correlation】来检验和测量非独立性。

要解决由非独立性造成的问题，我们一般要用偶体，而非单体作为分析的单位。在有些情况中，可用偶体的两个成员之间的和或差作为分析的变量来进行分析。尽管非独立性可以被看成是一个问题，但是它也给我们带来了研究各种偶体现象的机会。例如互惠、补偿、相似和同步意味着偶体成员回应中的非独立性。诸如这样的非独立性可能是由非随机配对（如选择性交配）、共命运和相互影响造成的。

在偶体分析中，偶尔也会为每个偶体创建一个指数。例如将相关系数作为成员的闲暇偏好的相似性指数。尽管偶体指数可能有很大的信息量，但是在对它们进行解释时也存在着若干误区（Cronbach，1955）。

在偶体分析中，研究一个人的特性是如何影响自己和他的伙伴的行为是很有用处的。这样，对每个人而言，都存在一种对他自己和伙伴的特性进行回归的结果。举一个例子，我们可以考察一个人和他的室友的抑郁程度如何影响他对室友的满意度。凯西和肯尼（Kashy & Kenny，2000）把这一模型称为客体互倚模型（actor-partner interdependence model）。有时，我们也可以用反馈回路和工具变量来估计相互影响模型（Duncan，Haller，& Portes，1975）。

有时，有的偶体数据将每个人和若干个他人配对。例如，每个人可以给她的女学生联谊会中的每一个其他的人打分。这样的循环数据（round-robin data）需要使用社会关系模型【Social Relations Model】。

——David A. Kenny

（沈崇麟译校）

## 参考文献


Cronbach，L. J.（1955）. Processes affecting scores on "understanding of others" and "assumed similarity." *Psychological Bulletin*，*52*，177-193.

Duncan，O. D.，Haller，A.，& Portes，A.（1968）. Peer influence on aspiration：A reinterpretation. *American Journal of Sociology*，*75*，119-137.

Kashy，D. A.，& Kenny，D. A.（2000）. The analysis of data from dyads and groups. In H. T. Reis & C. M. Judd（Eds.），*Handbook of research methods in social and personality psychology*（pp.451-477）. New York：Cambridge University Press.


# 动态建模（Dynamic Modeling）

参见模拟【Simulation】。

# E

## 区位谬误(Ecological Fallacy)

在19世纪的欧洲,许多国家特别是在新教国家的自杀率居高不下。爱米尔·涂尔干推断自杀是由新教的社会条件所致(Le suicide,1897)。这是一个"区位推论"——一个从群体行为的数据得出的关于个体行为的结论。

当然,新教国家除了宗教之外,在许多方面和天主教国家都有所不同("混杂"问题)。不仅如此,涂尔干的数据并未将个人的自杀和任何宗教信仰联系起来:他针对的是地理区域内的国家而不是教派。混杂的问题必须用任何观察的数据来处理,而第二个问题则是针对区位研究的:推定的原因和效应是对群体而非个体量度的。不仅如此,分析的兴趣是一种组合——对涂尔干而言是宗教,而可资利用的数据却是一种完全不同的组合(地理)。

如果没有混杂,那么期望的群体和个体之间的差异就是"聚集偏倚(aggregation bias)";一般来讲,这种差异部分归结于混杂,部分归结于聚集偏倚。"区位谬误"源于认为在群体间观察到的关系,必然也存在于个体之间这样一种思想:如果有较多新教徒的国家倾向于有较高的自杀率,那么新教徒就必定更可能自杀。

区位研究虽然可以提供有用的线索,但一般来讲,有关个体的结论只能从群体的数据中得到微弱的支持。问题的原因正是混杂和聚集偏倚。正如罗宾逊(Robinson,1950)所言,个体层次的关系可能被聚合所颠倒。这就是相关系数的"辛普森悖论"。有几种统计方法被建议用来将个人层次的性状从群体层次性状中分解出来。这些方法包括"区位回归(ecological regression)"和"交叉层"或"层级"回归等。然而每种方法本身都有相当强的性状假设。但这些假设一旦被清楚陈述,似乎都是难以置信的。例如,区位回归做的"不变性(constancy assumption)"假定。根据这一假定,像涂尔干这样的个体的性状取决于宗教信仰的派别,而不是地理位置的应用,不仅全欧洲的新教徒必须都有相似的自杀习性,而且天主教徒也必须都是同质的。

一般来讲,模型中的识别假定是不能用数据验证的。此外,在有个体层次的数据可用来进行检验时,便可将估计值与现实作比较,但模型追踪的记录充其量也只不过是混合的而已。其他各个方面的各种讨论,请参见下面的参考文献。

——David A. Freedman

(沈崇麟译校)

### 参考文献


Firebaugh, G. (2001). Ecological fallacy. *International encyclopedia for the social and behavioral*

*sciences* ( Vol. 6, pp. 4023-4026 ). Oxford, UK: Pergamon.

Freedman, D. A. ( 2001 ). Ecological inference and the ecological fallacy. *International encyclopedia for the social and behavioral sciences* ( Vol. 6, pp.4027-4030). Oxford, UK: Pergamon.

Freedman, D. A., Klein, S. P., Ostland, M., & Roberts, M. R. ( 1998 ). Review of "A solution to the ecological inference problem." *Journal of the American Statistical Association*, *93*, 1518-1522. (Discussion appears in Vol.94, pp.352-357.)

Goodman, L. ( 1953 ). Ecological regression and the behavior of individuals. *American Sociological Review*, *18*, 663-664.

Grofman, B., & Davidson, C. ( 1992 ). *Controversies in minority voting: The voting rights act in perspective.* Washington, DC: Brookings Institution.

King, G. ( 1997 ). *A solution to the ecological inference problem.* Princeton, NJ: Princeton University Press.

Robinson, W. S. ( 1950 ). Ecological correlations and the behavior of individuals. *American Sociological Review*, *15*, 351-357.

## 区位效度(Ecological Validity)

区位效度也称为生态效度,是指在一种环境(经常指一种实验装置)中研究的行为指示,可在多大程度上作为(或概括为)在其他一系列环境(常被称为"日常"或"自然")中的个人的认知过程特性。

有关区位效度的最初讨论出现在1940年代两位美国学者埃贡·布伦瑞克(Brunswik,1943)和库尔特·卢因(Lewin,1943)的心理学研究中。他们二人于1930年代从德国移民到美国。沿着这一传统,发展心理学家尤里·布朗芬布伦纳提出了正当的区位研究必须满足的三个条件:(a)保持被设计用于调查的现实生活环境的完整性;(b)忠实于应试者被征召的那个较大的社会和文化情境;(c)实验的操作和结果是能演示的,"以一种与明确和隐含在实验设计中的概念定义一致的方式为参与者所感知"(Bronfenbrenner,1979,p.35)。在一些类似的术语中,奈塞尔(Neisser,1976)注意到:一方面,在空间、时间和综合关系之间存在明显的不连续性;另一方面,基于实验的研究的物体和事件的特性是认知心理学的一个根本的不足。

许多区位效度的讨论都假定在实验任务中发现和直接观察到的途径在非实验设置中也会发生(或不发生)(如Simon,1976)。然而,当代研究的各种各样的发现中表明,建立区位效度的要求已经成为社会科学家沉重的分析负担(参见Cole,1996,Chap.8,9,对相关问题的扩展处理)。一旦我们走出实验室,到现实世界去找类似的事物,确定任务的能力就会明显变弱。无法定义分析者的任务的参数或无法保证被发现的任务也是应试者的任务都对任务造成了损害。

于是问题出现了,正如有学者(Valsiner & Benigni,1986)指出的那样,因为标准化的认知实验程序意味着所体现的系统是封闭的分析系统。其目的是试图建立任务等价,以便进行超越实验环境的概括,实际上是将一个更开放的行为系统地施加于一个封闭系统。迄今为止,社会科学家的封闭系统,始终未能有效地捕捉到假定要进行建模的开放系统的元素,因此,实验结果系统地歪曲了它们被从中导出的生命过程。

——Michael Cole

(沈崇麟译校)

## 参考文献

Bronfenbrenner, U. (1979). *The ecology of human development*. Cambridge, MA: Harvard University Press.

Brunswik, E. (1943). Organismic achievement and environmental probability. *The Psychological Review*, *50*, 255-272.

Cole, M. (1996). *Cultural psychology: A once and future discipline*. Cambridge, MA: Harvard University Press.

Lewin, K. (1943). Defining the "field at a given time." *Psychological Review*, *50*, 292-310.

Neisser, U. (1976). *Cognition and reality: Principles and implications of cognitive psychology*. San Francisco: W. H. Freeman.

Simon, H. A. (1976). Discussion: Cognition and social behavior. In J. S. Carroll & J. W. Payne (Eds.), *Cognition and social behavior*. Hillsdale, NJ: Erlbaum.

Valsiner, J., & Benigni, L. (1986). Naturalistic research and ecological thinking in the study of child development. *Developmental Review*, *6*(3), 203-223.

# 计量经济学(Econometrics)

计量经济学在字面上的意思是"经济计量"。尽管计量是计量经济学的重要部分，但计量经济学的范围要比它广泛得多，从下面摘录的文字我们便可看到这一点。

"计量经济学是对经济学作用的一种深入观察的结果，它将数理统计应用于经济数据，将经验支持施予由数理经济学构建的模型，进而得到数值结果"(Tintner, 1968, p.74)。

"计量经济学也许可以定义为一种将经济学理论工具、数学和统计推论应用于经济现象分析的社会科学"(Tintner, 1968, p.74)。

"计量经济学研究方法的目标基本上在于用理论和统计推论技术作为桥梁，将理论和实际的计量结合起来"(Haavelmo, 1944, p.iii)。

诚如前述定义，计量经济学是经济学理论、数理经济学、经济统计学和数理统计的合成。它不是这些学科中的某一个的附庸，现在计量经济学在大多数商学和经济学的系科中都是单独教授的，其原因我们将在下面进行阐述。

经济学理论作出陈述或假设，在实质上它们通常都是定性的。例如，微观经济学理论认为，在其他事情相同时(著名的其他条件不变[Ceteris Paribus]条款)，一种商品价格的下降，则那种商品的需求量上升可期。因此，经济学理论假定商品的价格和需求之间存在一种负或反关系。然而这个理论本身并未给我们提供二者之间的关系的数值计量。这就是说，它并未告诉我们作为那一商品价格的变化结果，需求量将会上升或下降多少？而提供这样的数值估计则正是计量经济学要做的工作。换言之，计量经济学将经验内容给予了经济学理论。有时，计量经济学可以通过明确地考虑归属于其他条件不变条款的变量，对现存的理论加以改进。定性变量，例如性别、宗教和种族被加进了劳动力市场研究，用来解释劳动力的参与(如工作与否的决策)。在传统上，诸如计时工资和教育程度这样的变量都被视为劳动力参与问题的主要决定性因素。

数理经济学关心的主要问题是在不考虑计量或理论的经验验证情况下，以数学的

形式(方程)表达经济学理论。虽然计量经济学也使用由数量经济学家提供的数学方程,但它将这些方程表达为可做经验检验的方式。这种将数学方程式变成计量经济学的方程(如随机方程)需要高度的独创性和实际技能。

经济统计学主要关心如何用图表的形式收集、处理和展示经济数据。这些工作都由经济统计学完成。从事这些工作的人,主要负责诸如国内生产总值(GDP)、就业、失业或价格这样的变量的数据收集。经济统计学家不再做进一步的工作,除了关心如何用收集的数据检验经济学理论之外。而有些经济统计学家,则做了进一步的工作,于是他们就变成了计量经济学家。

虽然数理统计学提供了许多用于贸易的工具,但计量经济学家需要观察大多数经济数据,即那些并非作为控制的实验结果产生的数据的独特性质的特殊方法。计量经济学家,与普遍依靠不能直接控制的数据的气象学家颇为相似。正如斯帕诺斯(Spanos,1999)观察到的:

“在计量经济学中,模型制造者常常面对的是与实验数据截然不同的观察数据。这包含着计量经济学建模中两个重要的含义。第一,建模者需要比分析经验数据更多的各种不同的技巧……。第二,数据收集者和数据分析者的分离,要求建模者本身非常熟悉正在研究的数据的性质和结构。”(p.21)

## 计量经济学的方法学

计量经济学家是如何分析一个经济学问题的呢?也就是说,他们的方法学是什么呢?一般来讲,计量经济学的方法学有三种路数:经典路数、Leamer(Leamer,1978)路数和 Hendry(Hendry,1995)路数。

所有这些计量经济学的方法学路数的讨论都非三言两语能说清楚的。经典方法学的主要特点将在这里予以阐述,不仅是因为它重要的历史地位,更主要的是这种方法学仍然主导着许多领域的经验工作。在讨论了这一方法学之后,我们再讨论其他两种方法。

## 经典方法学

经典方法学的实际进程如下:

经济学理论→理论的数学模型→理论的计量经济学模型→数据→估计计量经济学模型→假设检验→预测或预报→将模型用于控制或政策的目的。

每一个箭头都表示一个连续的或逐步的过程。

为了更好地阐明经典方法学,我们来看一下著名的凯恩斯消费理论。

### 理论或假设陈述

凯恩斯(Keynes,1936)写道:

“基本的心理学法则……是男人(和女人)都是有打算的,作为一种规则和平均而言。他们的消费会因为收入的提高而提高,但消费的提高往往不如收入的提高那么多。”(p.36)

简言之,凯恩斯假定边际消费倾向(marginal propensity to consume, MPC),即一个单位(比如说,一美元)的收入变化的消费变化率,大于零但小于1。

### 消费数学模型的设定

虽然凯恩斯假定消费和收入之间的正关系,但是并没有设定二者之间的精确关系。为了简便起见,数理经济学家可能会提出下面的凯恩斯消费函数形式:

$$Y = \beta_1 + \beta_2 X;\ 0 < \beta_2 < 1 \qquad (1)$$

式中的 $Y$ 是消费支出,而 $X$ 则是收入,$\beta_1$ 和 $\beta_2$ 是模型参数,分别是截距和斜率系数。

用斜率系数表示 MPC。

### 消费计量经济学模型的设定

消费函数的数学模型在方程(1)中给出,但是它限制计量经济学家的使用,因为它假定在消费支出和收入之间存在着确切或确定的关系。但经济变量之间的关系一般都是不确切的。为了能使用不确切关系,计量经济学家可能会对消费函数模型进行修正,将它改写成:

$$Y = \beta_1 + \beta_2 X + u \qquad (2)$$

式中的$\mu$称为干扰或误差【Error】项,是随机变量【Random Variable】,有很好定义的性质,误差项$u$可能代表所有那些影响消费,但在方程(2)中没有清楚地解释。方程(2)是计量经济学模型的一个例子。更严格地讲,这是一个线性回归模型【Linear Regression Model】的例子。

### 获取数据

为了估计方程(2),也就是求$\beta_1$和$\beta_2$的数字估计值,而在目前的语境中,它们称为回归系数——我们需要的数据。为了说明这一问题,我们得到了美国 1982—1996 年的$Y$(个人消费支出)和$X$(GDP)的数据,该数据在 1992 年是以亿美元计。

### 估计计量经济学模型

用最普通小二乘【Ordinary Least Squares, OLS】,并根据美国的数据,我们得到了下面的回归结果:

$$\begin{aligned} &\hat{Y}_t = -184.078\,0 + 0.706\,4X_t \\ &SE:(46.261\,9)(0.007\,8) \qquad (3) \\ &R^2 = 0.998\,4 \end{aligned}$$

式中$Y_t$上的“^”的意思是,它是消费支出的均值或平均数的估计值,而括弧中的数字是估计的标准误差。下标$t$表示时间。[经典线性回归对误差项$u$做了某种假定,即它有零均值、常方差,且误差是不相关的。对于假设检验,它经常假定误差项是正态分布的。如果正态性被违反了,那么我们可以使用最大似然估计【Maximum Likelihood Estimation, ML】来估计回归模型的参数。$\beta_1$和$\beta_2$的 ML 估计值与 OLS 估计的相同]。

### 假设检验

凯恩斯断言消费的边际倾向是正的,但是小于 1。在方程(3)给出的估计回归模型中,MPC 是 0.706 4。它是否在统计上不同于1? 换句话说,我们希望检验一个 MPC 是 1 的虚无假设。那么,我们应该如何来检验这一假设呢?

基于经典线性回归【Classical Linear Regression】模型给定的假定,我们可以用$t$检验【$t$-Test】来检验这一假设,用目前这一例子,我们可以得到

$$t = \frac{0.706\,4 - 1}{0.007\,8} = -3.764\,1 \qquad (4)$$

因为数据中有 15 个观察,所以这一$t$值有 13 个自由度【Degrees of Freedom】。因此,这一$t$值在 1%的水平上是显著的(实际的$p$值要低得多)。因此,我们拒绝真 MPC 是 1 的假设或许可证实凯恩斯的假设。

### 预测与预报

假定计量经济学模型是现实的一个相当好的近似,我们便可以根据预测变量$X$的已知或期望值用它来预测未来的值。假设我们打算预报 1997 年的平均消费支出数。

假设 1997 年 GDP 的值是 7 269.8 亿美元（它是一个实际的值）。在方程(3)中用这个 GDP 的值,得到预报的平均消费支出约为 4 951.31 亿美元。读者可自行效验这一数字。

将模型用于控制和政策的目的

假设政府相信消费支出约为 4 900 亿美元能使失业保持在目前的水平。那么什么样的 GDP 水平将保证消费支出的目标水平呢？假定在方程(3)中估计的消费支出回归相当好,那么我们只需要将 4 900 亿美元放在方程的左边,并求出 $X$ 即可。简单的四则运算便可求得保持目前的失业水平的支出水平为 7 200 亿美元。在这一练习中, $Y$ 是目标变量,而 $X$ 则是控制变量。

前述的 8 个步骤是简约的经典方法学。对于这种方法学存在一些批评意见。其中两种意见将在下面进行讨论。

## Hendry 自上而下方法学

在亨德里(Hendry,1995)看来,经典方法学过于静态。这就是说,它没有将随时间的变化考虑在内。换句话说,他将在一种动态的环境中来开发计量经济学模型。他将使他的计量经济学模型先尽可能地全面,然后再经过大量诊断,将它消减成比较小的模型。因为亨德里最初的大部分工作是在伦敦经济学院(London School of Economics, LSE)做的,因此他的方法学也称为 LSE 方法学。

LSE 的起点是经济学理论。该理论假定经济变量,比如说,$Y$(“永久”消费)和 $X$(“永久”收入)之间存在长期或均衡的关系,“永久”这一词表明家庭认为在将来可能保持的水平。这一关系可概括为

$$Y_t = \alpha X_t \qquad (5)$$

亨德里和他的同事之所以在 LSE 指数中让观察带有时间下标 $t$,是因为他们的方法学被开发出来主要用于时间序列【Time-Series】的经济数据。

当然,在方程(5)中假定的长期关系取了达到的时间,因此,LES 方法学建议采用下面的动态程序形式,以达到方程(5)

$$Y_t = \beta_0 X_t + \beta_1 X_{t-1} + \cdots + \beta_m X_{t-m} + \delta_1 Y_{t-1} + \delta_2 Y_{t-2} + \cdots + \delta_l Y_{t-l} + u_t \qquad (6)$$

在这一模型中,$Y$ 不仅有赖于当前和滞后的 $X$ 值,而且也有赖于它自己的滞后值,滞后值由滞后下标记号表示。当然更多的 $X$ 变量和它们的滞后值可以被引进。

正如方程(6)所示,我们将在时间 $t$ 的 $Y$,对在时间 $t$,$(t-1)$,$(t-2)$,…,$(t-m)$的 $X$ 的值及 $Y$ 在时间$(t-1)$,$(t-2)$,…,$(t-l)$的滞后值做回归,滞后项 $m$ 和 $l$ 的选择则是一个经验问题。

方程(6)中的模型是自回归分布滞后(autoregressive distributed lag, ADL)的一个例子。它之所以是自回归的,是因为我们将当前的 $Y$ 值,对它前面的滞后值做了回归,而它是被分布的,因为 $X$ 的影响在时间上是延续的。这种 ADL 模型也被称为动态模型,因为它们很清楚地考虑了时间延续上的 $Y$ 的性状。方程(6)一旦被估计,我们就可以用恰当的代数运算返回到长期关系。

## Leamer 方法学

利默是一位经典计量经济学的严厉批评家,他将其称为平均经济回归(average economic regression, AER)。在他看来,在

非经验收集的数据中，AER 策略是有问题的。对于他和其他批评家来讲，模型一旦设定，估计它的参数和进行假设检验便是微不足道的，而一开始确定什么样的模型才是恰当的这一任务才是真正需要功夫的。这一任务就是特殊计量学（specimetrics）的主题。利默（Leamer，1978）认为：

“特殊计量学描述一个引导研究者选择某一个，而非另一个模型设定的过程，此外，它还试图在数据的产生机制不太明确的时候，确定可能从一个数据集中做出的适当的推论。”（p.v）

简言之，利默和他的追随者认为，我们必须万分小心地注意特殊计量，即选择适当的模型。只有在模型选择了之后，我们才可以论及经典或 AER 方法学。

利默和 AER 方法学除了这一基本不同之外，利默还是贝叶斯方法的坚定支持者。计量经济学建模的经典和亨德里路数是以经典统计学为基础的。在贝叶斯路数中，比如说在回归模型中，方程（1）中像 $\beta_1$ 和 $\beta_2$ 这样的“参数”被看成是“随机的”，每个都有着某种分布（在贝叶斯的话语中称为先验）。先验一般是根据诸如先前的研究和对一个现象的理解这样的因素主观选择的。如果我们的手中握有一组数据，我们先估计 $\beta_1$ 和 $\beta_2$，然后使用样本估计值和先验得到的结果称为后验（概率）分布。因此，在贝叶斯方法中，$\beta_1$ 和 $\beta_2$ 一部分是从数据估计而来的，而一部分则是由有关这些系数的先前的信息估计而来。这给了我们这些系数的后验（概率）分布。在下一轮分析中，这些后验分布变成了先验分布。贝叶斯方法是一种“用做来学的”程序。在有更多数据变成可资利用时，研究者可用它们来修正先前有关这些系数的知识。

另一方面，在经典回归程序中，$\beta_1$ 和 $\beta_2$ 被看成是定数，尽管它们的值是未知的。假如有一组数据，我们可以用 OLS 或最大似然估计参数，并进行假设检验。假如我们又有了一组新的数据，我们还是重新开始，并没有将先前的样本结果考虑进来。因此，在这里，这两种路数存在着根本的差别。

——Damodar N. Gujarati

（沈崇麟译校）

## 参考文献


Goldberger, A. S. (1964). *Econometric theory*. New York: John Wiley & Sons.

Gujarati, D. (2002). *Basic econometrics* (4th ed.). New York: McGraw-Hill.

Haavelmo, T. (1944). The probability approach in econometrics. *Econometrica*, *12* (Suppl.), preface, p. iii.

Hendry, D. F. (1995). *Dynamic econometrics*. New York: Oxford University Press.

Keynes, J. M. (1936). *The general theory of employment, interest, and money*. New York: Harcourt Brace Jovanovich.

Leamer, E. E. (1978). *Specification searches: Ad hoc inference with nonexperimental data*. New York: John Wiley & Sons.

Mittelhammer, R. C., Judge, G. G., & Miller, D. J. (2000). *Econometric foundations*. New York: Cambridge University Press.

Spanos, A. (1999). *Probability and statistical inference: Econometric modeling with observational data*. Cambridge, UK: Cambridge University Press.

Tintner, G. (1968). *Methodology and mathematical economics and econometrics*. Chicago: University of Chicago Press.

# 效应量(Effect Size)

效应或关系量级的标准化数字指数【Index】称为效应量。在上述定义中"量级"和"标准化"这两个词尤其重要,因而值得作进一步的解释。效应的数量级是对社会科学的发现作出适当解释的关键之所在。与效应量的考量有别的做法是,仅依据统计显著性检验【Significance Testing】的结果做结论。因为统计显著性受到样本量的影响是如此之大,致使那些微弱的,或许是微不足道的效应由于大样本而在统计上仍然是显著的。另一方面,如果样本量很小,某种很强的效应则可能在统计显著性检验时未被探测到。因此,在使用统计显著性检验之外,最好能把我们的结论也以效应量为依据,因为它被设计成了一种独立于样本量的量。

实际上,效应量是以标准化形式而非原始分数形式陈述的,所以单位是很重要的,在社会科学中尤其如此。我们的许多量度,其量度量表都有一定的任意性(如李克特量表【Likert Scales】)。因为量表是任意的,所以它们没有固有的含义,含义可能随研究的变化而变化。假设在实行给定的工作场所干预之后,雇员的工作满意度较未实施干预之前平均高 0.5 个单位。如果仅就工作满意度可用无数个量表测量这一事实而言,那么这一信息的接受者,除了高 0.5 个单位外没有任何其他意义。0.5 这个值可能是一个很大的变化,也可能是几乎没有变化。而在另一方面,如果这个结果以"后干预的工作满意度较干预之前高 0.5 个标准差【Standard Deviations】"这样一种形式报告的话,那么下起结论来就要容易得多。不论是什么标准差(即不论是什么量表),后干预的满意度都高半个标准差。一个试图复制这一发现的研究可以使用一个不同量表的满意度量度,但是如果它的结果是以标准差为单位报告的,那么将可直接与那些任何诸如这样的研究进行比较。正因如此,效应量通常都是标准化的。

尽管有着许多不同形式的效应量指数,但值得在这里一提的只有少数几个。对相关数据而言,使用最为普遍的效应量是相关【Correlation】系数的平方。在其他类型中,系数本身的类型可能有简单相关、多重相关或偏相关。系数的平方则尤其便利,因为它反映了一个变量中为另一个变量【Variable】所解释的方差【Variance】的百分比。

对那些求均值差【Difference of Means】的方法而言,例如,那些与方差设计【Analysis of Variance】有关联的方法,有着各种各样可资利用的指数。其中最为普遍的是 Cohen $d$(Cohen's $d$)统计量,它以各种方式来定义,其中最为普遍的是两个均值之间的差除以合并标准差,即 $s_{合并}$,它是组内方差和样本量的函数。

$$s_{合并} = \left\{\frac{(N_1 - 1)s_1^2 + (N_2 - 1)s_2^2}{N_1 + N_2 - 2}\right\}^{0.5}$$

如果因变量是以单位方差来量度的话,那么求得的 $d$ 值直接就是均值差。$d$ 还可解释为一组中那些小于另一组的平均分数的分数的比例。

应当注意的是,这里给出的两个效应指数的例子都反映了效应的量级,且都已被标准化。

——Jose M. Cortina

(沈崇麟译校)

# 效应编码(Effects Coding)

效应编码是一种为 $j-1$ 个虚拟变量【Dummy Variables】编码,以充分代表包含在 $j$ 个类别的分类中的信息。效应编码的虚拟变量系数,允许研究者将指定类别的预测值和总样本的“平均”值作比较。这种“平均值”是组均值的未加权均值。未加权均值的计算方法是 $[\sum_{j=1}^{J}\bar{X}_j]\div J$。只有在子组大小相等时,组均值的未加权均值才等于总均值(Hardy,1993)。因此,回归系数对特定的子组的均值和所有子组均值的未加权均值之间的差进行了量化。

用一个例子来做进一步说明。我们来考量一下婚姻状况(已婚、分居、离婚、鳏寡、从未结婚)和计时工资(以美元计)之间的关系。在这个例子中,婚姻状况有五个($J$)子组,由四个($J-1$)虚拟变量代表。虚拟变量的效应编码列在了表 1 中。我们选择已婚作为省略的类(Hardy,1993),它在所有四个虚拟变量上的编码都是-1,而每个虚拟变量将其余四个子组中的一个的观察编码为“1”,而将其余三个子组的观察编码为“0”。例如,从未结婚效应(EFFNM)(依据效应编码的虚拟变量),将从未结婚编码为“1”,已婚的编码为“-1”,而其余的子组(分居、鳏寡和离婚)编码为“0”。

表 1

| 子组 | 效应编码虚拟变量实例 | | | | | |
|---|---|---|---|---|---|---|
| | 从未结婚效应 | 分居效应 | 鳏寡效应 | 离婚效应 | 均值(以美元计) | 个案数 $N$ |
| 已婚 | -1 | -1 | -1 | -1 | 15.33 | 1 661 |
| 从未结婚 | 1 | 0 | 0 | 0 | 12.95 | 613 |
| 分居 | 0 | 1 | 0 | 0 | 11.64 | 146 |
| 鳏寡 | 0 | 0 | 1 | 0 | 12.94 | 390 |
| 离婚 | 0 | 0 | 0 | 1 | 11.93 | 14 |
| 所有事件 | | | | | 14.26 | 2 773 |

我们来看一下效应虚拟变量的描述性统计值和零阶相关。每个效应编码的虚拟变量的均值报告的是参照组(编码为-1)和编码为 1 的子组之间大小的差异,如($n_j-n_{ref}$)/$N$。负号表明参照组中的观察多于子组 $j$;反之,正的均值则表明子组 $j$ 比参照组有更多的观察,而均值的数量则表明了相对于总样本量的组的大小差异或多出的个案比例。

在效应虚拟变量被用作普通最小二乘回归的自变量时,常数报告的是子组均值的未加权均值。系数报告的是子组均值与这一参照点的偏差。样本的回归结果已列在了表 2 中,它是计时工资对婚姻状况的效应虚拟变量的回归。表 2 中的系数是未标准化的。

表 2

| 模型 | $B$ | 标准误差 | $t$ | 显著性 |
|---|---|---|---|---|
| 常数 | 12.96 | 0.471 | 27.527 | 0.000 |
| 从未结婚效应 | −0.005 | 0.536 | −0.010 | 0.992 |
| 分居效应 | −1.318 | 0.704 | −1.871 | 0.061 |
| 离婚效应 | −0.022 | 0.569 | −0.038 | 0.970 |
| 鳏寡效应 | −1.027 | 1.755 | −0.585 | 0.559 |

全部观察（$N = 2\ 773$）的总均值是 14.26，而子组的未加权均值是 12.96，可由将 5 个均值相加再除以 5 求得。这一数字即表 2 中的常数项(constant)。在评估 $t$ 检验时，我们看到鳏寡、离婚、分居和未婚的子组均值与这一值并无显著的差别。例如，分居(EFFS)的系数−1.318 表明，分居的人的计时工资比 12.96 美元的未加权均值少1.32 美元，为 11.64 美元。尽管每小时的差异大于 1 美元似乎是相当大的，但是标准误差表明这个工资差距的量度是不够可靠的(例如，95%的置信区间为−1.70～1.60)。然而 $F$ 检验却表明在优于 0.001 的阿尔法水平上，我们可以拒绝虚无假设(假设不同婚姻状态的计时工资是没有差别的)。因为那些已婚的人的小时工资为 15.33 美元[−1×(−0.005−1.318−0.022−1.027)+12.960=15.33]，它既显著不同于 12.96 美元，也显著不同于所有其他四个子组的平均计时工资。

——Melissa A. Hardy

Chardie L. Baird

（沈崇麟译校）

＊也可参见编码【Coding】。

## 参考文献


Cohen, J., & Cohen, P. (1983). *Applied multiple regression* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Hardy, M. A. (1993). *Regression with dummy variables*. Newbury Park, CA: Sage.


# 效应系数(Effects Coefficient)

一个变量在路径模型中的总效应，包括直接的和间接的，称为效应系数。假设下面的递归路径【Recursive Path】模型：

$$Y = a + bX + e$$

$$Z = c + dX + fY + u$$

变量 $X$ 在 $Z$ 上有用 $d$ 估计的直接效应和通过在 $Y$ 上的影响，用 $b×f$ 或 $bf$ 估计的间接效应。因此，$X$ 在 $Z$ 上的效应系数是 $d+bf$。效应系数使分析师能在一个因果【Causal】系统中，同时考虑间接和直接效应的，进而对一个变量的影响，按它在一个因果系统中的运行方式加以概括。不过，如果那个变量并非真正外生的话，那么效应系数的计算可能就会有问题。

——Michael S. Lewis-Beck

（沈崇麟译校）

＊也可参见路径分析【Path Analysis】。

## 效率(Efficiency)

一个有效的估计量有一个有最小方差【Variance】的概率【Probability】分布。例如,$W$ 和 $Z$ 都是参数的估计量【Estimators】,而它们的方差分别为 var($W$)和 var($Z$)。如果 var($W$)< var($Z$),那么 $W$ 是一个比 $Z$ 更有效的 $\theta$ 估计量。而这个效率常常被视作一种相对效率,因为效率需要在两个或更多个估计量之间进行比较。效率是估计量的一个合乎要求的性质,因为有着较小的方差的估计量,在任何给定的试图产生真 $\theta$ 值时都有较大的机会,只要这个估计值是无偏的【Unbiased】。

为了对效率进行直观的阐述,我们来看一下两个玩飞镖的人。我们的目标是尽可能多地击中牛的眼睛。每一个人代表一个不同的估计量,而牛的眼睛则是每个估计量的概率分布的均值。玩家 1 投了 100 次,飞镖都落在了牛眼的 5 英寸的圆圈内。玩家 2 也投了 100 次,但飞镖都落在牛眼的一个 12 英寸的圆圈内。玩家 1 比玩家 2 更有效,因为他的飞镖更接近均值(牛的眼睛)。玩家 1 也比玩家 2 有着更高的击中牛眼的概率,因为玩家 1 的投标能力的方差更小,只要两个玩家的投掷是无偏的(这就是说,他们可能站在了牛眼的同一侧,且距牛眼的距离也相同)。

一个有效的估计量并非总是能产生一个等于真的 $\theta$ 值的估计值,因为效率是那些估计值从中抽取的概率分布的一种性质。这就意味着,在进行反复抽样时,即使是一个有效的估计量,也有可能会产生不等于真的 $\theta$ 的估计值。然而,若使用了一个有效的估计量,那么我们就有比较大的把握确定,估计值是接近真的 $\theta$ 值的。此外,如果估计量是有偏的,那么一个有效的估计量将无法产生等于它的真值的 $\theta$ 的估计值。

因为估计量是如此之多,以致在确定估计量的效率时,必须对我们比较的估计值的种类有所限制(Kennedy,1993)。首先,我们只看那些无偏的估计量。一般来讲,无偏估计量产生的 $\theta$ 的估计值将等于 $\theta$ 的真值。通过将我们的比较限制在只是在无偏统计量的范围,我们可以只比较那些有相同的均值分布和产生的通常是等于参数的真值的参数估计值。在很多时候,我们难以在几个无偏统计量中确定究竟哪一个才是效率最高的时候,还可加上另外一个限制,把我们的比较仅限于那些线性估计量。有着最小方差的线性无偏估计量代表一个特殊的估计量族,称为最佳线性无偏估计量【Best Linear Unbiased Estimator, BLUE】。

——Heather L. Ondercin

(沈崇麟译校)

### 参考文献

Kennedy, P. (1993). *A guide to econometrics* (3rd ed.). Cambridge, MA: MIT Press.

## 特征值(Eigenvalues)

特征值也称为特征根(characteristic roots)或本征根(latent roots),是一种与矩阵【Matrix】方程有关的一组特殊的标量。在一些统计方法中,包括因子分析【Factor Analysis】和主成分分析【Principal Components Analysis】,它们都是非常重要的数学概念。

令 $\mathbf{A}$ 为一个 $k \times k$ 的复数方阵,如果存在一个非零的 $k \times 1$ 的列向量 $\mathbf{X}$,那么,一个

属于复数集合的标量 $\lambda$,便可以说是 **A** 的一个特征值,致使

$$\mathbf{AX} = \lambda\mathbf{X}$$

列向量 **X** 称为 **A** 的特征向量,因为它被置于 **A** 的右边,故而称为“右特征向量”。一个行特征向量被置于 **A** 的左方,称为“左特征向量”。每个特征值都和一对特征向量——左和右特征向量关联。将 **A** 分解成特征值和特征向量称为特征分解。上述方程也可更为简洁地表示为:

$$(\mathbf{A} - \lambda\mathbf{I})\mathbf{X} = 0$$

式中的 **I** 是单位矩阵。在一个特征值与其他所有的特征值都不同时,它的特征向量是唯一的。我们用一个例子来说明特征值的应用。来自主成分分析的一个维度的特征值量度了拟合优度,并给出了在原始变量中的方差为主成分所解释的比例。

——Tim Futing Liao

(沈崇麟译校)

* 也可参见因子分析【Factor Analysis】。

## 特征向量(Eigenvector)

参见特征值【Eigenvalues】、因子分析【Factor Analysis】。

## 详析(拉扎斯菲尔德法)(Elaboration[Lazarsfeld's Method of])

从最基本的形式来看,详析法就是从一个二维的列表开始,先展示两个变量之间的关系,然后再引入第三个变量进行详析和说明。表 1[本例系受拉扎斯菲尔德(Lazarsfeld,1955)和海曼(Hyman,1955)的基础文本的启发]是一个处理二分变量 *X*-性别(1=女性,2=男性)和 *Y*-去年的车祸(1=有,2=无)之间的(虚假)关系的例子。这个表是用百分数来表示的,括号里的数字是绝对数。

**表 1**

| *Y*-车祸 | *X*-性别 | | 总计 |
|---|---|---|---|
| | 1=男性 | 2=女性 | |
| 1=有 | 38%<br>(100) | 19%<br>(50) | 28%<br>(150) |
| 2=无 | 62%<br>(160) | 81%<br>(220) | 72%<br>(380) |
| 总计 | 100%<br>(260) | 100%<br>(270) | 100%<br>(530) |

女性似乎是比男性更好的司机:去年有 38%的男性司机发生了车祸,而女性司机只有 19%。造成这一差别的原因是什么呢?女性是否是更仔细、周到的司机?一个简单的解释可能是在过去的一年里,男性开车的里程大大多于女性。显然,一个人开车开得越多,发生交通事故的可能性就越大。这个详析法背后的逻辑是,如果后面这番解释是对的,那么比较每年开车里程数相同的男女司机,就会得到男性和女性的交通事故数没什么不同的结果。我们引入了第三个变量 *T*-里程(1=多,2=少),并调查在第三个变量的分类中(我们“控制 *T*”或者“我们保持 *T* 不变”来考察 *X* 和 *Y* 的关系)。表 2 显示的是第三个变量的引入。

表 2

| Y-车祸 | T-里程(1=多) X-性别 | | | T-里程(2=少) X-性别 | | |
|---|---|---|---|---|---|---|
| | 1=男性 | 2=女性 | 总计 | 1=男性 | 2=女性 | 总计 |
| 1=有 | 50% (95) | 50% (35) | 50% (130) | 7% (5) | 8% (15) | 7% (20) |
| 2=无 | 50% (95) | 50% (35) | 50% (130) | 93% (65) | 92% (185) | 93% (270) |
| 总计 | 100% (190) | 100% (70) | 100% (260) | 100% (70) | 100% (200) | 100% (270) |

在里程数多(*T*=1)的分表中,有交通事故的人大大多于里程数少(*T*=2)的分表,但在这些分表中(也即比较开车里程数相同的男女司机时)男性和女性就不再有差别了。拉扎斯菲尔德把这种类型的结果称为"阐释(interpretation)"。用箭头表示直接因果效应,这一结果的因果模式是 $X \rightarrow T \rightarrow Y$,意思是 *X* 是 *Y* 的一个原因,但间接通过了 *T*(因为在所有的社会环境下,女性都比男性开车要少,因此发生的车祸也较少)。如果在男性和女性之间还有其他的差异导致女性比男性发生的车祸少,那么即便在已经控制了 *T* 的分表中,男性和女性之间的百分数也会存在差异。

另一种类型的结果则是虚假的【Spuriousness】。对这种类型来说,与上一个例子一样,最初的关系【Association】也消失了,但现在 *X* 和 *T* 的因果次序却改变了:*X* 被视为 *T* 的一个结果而不是一个原因。因果模式是 $X \leftarrow T \rightarrow Y$。我们假定有一个表包含了 *X*-教育和 *Y*-保守主义,我们从中发现受教育较少的人比受教育更多的人更保守。同时,也可能对 *T*-年龄来说确实是人的年龄越大,就变得越保守;而且,年轻人平均来说受的教育更多。如果这里年龄要为教育和保守主义之间的关系负责,那么我们按上述方法控制年龄之后,原来在教育和保守主义之间的关系就会消失。从这个阐释结果中我们获悉 *X* 和 *Y* 之间是间接地有因果联系,而虚假关系则揭示了最初的相关【Correlation】的虚假,并证明这个关系并非因果的:年龄(*T*)既是受教育(*X*)的多少,也是保守主义(*Y*)的原因,但 *X* 绝不是 *Y* 的原因。

第三种类型的详析是详述或交互【Interaction】阐述。在这种类型中,分表中 *X* 和 *Y* 之间的关系是不同的。例如,在上面的第一个分表中(里程多的),可能会出现男性比女性引发的车祸更多,但当男女司机开车都不多(里程少)时,可能在男女司机之间就没差别了。交互阐述说明,在第三个变量的不同类别内,两个变量的(因果)关系如何不同。

详析法对分类变量所起的作用与多元回归分析【Multiple Regression Analysis】对连续变量【Continuous Variables】所起的作用十分相似。它的基本原理可以扩展到三个以上的变量和更为复杂的情况,尽管这样就需要更正式的方法了。这种正式方法在对数线性模型【Log-Linear Model】的环境中发展起来了。拉扎斯菲尔德的终极目标就是要用

“社会学的”数据发展出因果模型【Causal Modeling】的可靠方法，所谓“社会学的”数据，(在他看来)意思就是用非实验的调查设计获取的分类的【Categorical】数据。我们仍然可以在因果性的当代处理法中，如在珀尔的研究工作中找到他对因果分析的问题和可能性的深刻见解(Pearl，2000)。

——Jacques A. Hagenaars

(杨可译　沈崇麟校)

## 参考文献


Hagenaars, J.A.(1998).Categorical causal modeling: Latent class analysis and directed log-linear models with latent variables. *Sociological Methods and Research*, *26*, 436-486.

Hyman, H.(1955). *Survey design and analysis.* New York: Free Press.

Lazarsfeld, P. F.(1955). Interpretation of statistical relations as a research operation. In P. F. Lazarsfeld & M. Rosenberg(Eds.), *The language of social research*(pp.111-125).Glencoe, IL: Free Press.

Pearl, J.(2000). *Causality: Models, reasoning, and inference.* Cambridge, UK: Cambridge University Press.


# 弹性(Elasticity)

社会科学分析者经常对了解一个变量的变化引起另一个变量的变化这样的效应感兴趣。捕捉这样一种效应的方向和数量的一种有用的方法是弹性这一概念，它被定义为函数的弹性是在两个变量的值给定时，因变量的比例变化除以自变量的比例变化。这就是，

$$E=\frac{y\text{ 变化的百分数}}{x\text{ 变化的百分数}}=\frac{\frac{\Delta y}{y}\times 100}{\frac{\Delta x}{x}\times 100}=\frac{\Delta y}{\Delta x}\times\frac{x}{y} \quad (1)$$

式中变量前的符号 Δ 为变量的离散变化的取值，而 $y$ 和 $x$ 则分别为因变量【Dependent Variable】和自变量【Independent Variable】。

举例说明，假设我们对在特定饮食规定下的卡路里消耗对一个人的体重的效应的测量感兴趣。假定 $y$ 是以千克计的人的体重，而 $x$ 则是千卡计的每天卡路里的消耗。那么如果我们用绝对的变化，对前面的问题的答案将是比较随意的，因为它将受选择的单位的影响。这就是说，如果每天卡路里的消耗从 1 000 千卡增加到了 1 100 千卡，那么结果将是人们的体重从 60 千克增加到 63 千克，给我们的印象是，效应是相当小的。增加 100 千卡的卡路里消耗者只产生了 3 千克的体重增量。然而，假如我们把体重的计量单位从千克改为克，我们发现增加 100 千卡的卡路里将引起 3 000 克的体重增量。于是现在我们的印象是效应相当大。使用百分比变化的比率便可避免这一问题。无论体重是以千克计[(3 kg/60 kg)×100]，还是以克计[(3 000 g/60 000 g)×100]，给定的体重增量都是 5%。

有一个重要的问题提请大家予以注意：如果变量变化的取值不是太小，那么一般都取变化产生之前和之后变量的值之间的平均数。这时，弹性将取下面这样的表达式：

$$E=\frac{\Delta y}{\Delta x}\times\frac{\frac{x_1+x_2}{2}}{\frac{y_1+y_2}{2}}=\frac{\Delta y}{\Delta x}\times\frac{x_1+x_2}{y_1+y_2} \quad (2)$$

下标 1 和 2 分别表示变化产生之前和之后变量的取值。注意，如果变量变化的取值

不是太小，那么我们将会遇到方程（1）中的弹性在卡路里消耗增加和消耗减少时有所不同的问题。这就是说，在卡路里消耗从1 000千卡增加到1 100千卡时，弹性将是

$$\frac{3}{100} \times \frac{1\ 000}{60} = 0.5$$

当卡路里消耗从1 100千卡降到1 000千卡时，它将为

$$\frac{3}{100} \times \frac{1\ 100}{63} = 0.524$$

然而使用表达式（2），无论卡路里消耗是增加还是减少，我们得到的弹性估计值都等于0.512。

不仅如此，当Δ很小时（也就是无限小的增加），前面的公式将收敛于1，该计算要使用导数。也就是说

$$E = \frac{\partial y}{\partial x} \times \frac{x}{y} \tag{3}$$

弹性的概念在回归【Regression】分析中也非常重要。例如，在回归是一个像$y=\alpha+\beta x+\varepsilon$这样的线性回归时，方程（3）的使用给出的弹性等于$\beta(y/x)$。然而，在回归是对数或半对数回归时，有如下情形：

$$\ln y = \alpha + \beta_1 \ln x_1 + \beta_2 x_2 + \varepsilon$$

$$\text{使用}\begin{cases} E_{y,x_1} = \dfrac{\partial \ln y}{\partial \ln x} = \beta_1 & (4a) \\ E_{y,x_2} = \dfrac{\partial \ln y}{\partial x} \cdot x = \beta_2 x_2 & (4b) \end{cases}$$

和

$$y = \alpha + \beta \ln x + \varepsilon \quad \text{使用}$$

$$E_{y,x} = \frac{\partial y}{\partial \ln x} \cdot \frac{1}{y} = \beta \frac{1}{y} \tag{5}$$

我们用一个来自琳达海尔（Lindahl，2002，p.7）研究的例子来看一下它是如何运作的。该研究是一个有关1962年到1973年，瑞典的代际收入流动的研究。琳达海尔的研究的目的是得到一个机会平等程度的量度，平等程度则由代际收入流动来代表。在这里我们只对只有一个男孩的家庭这种情形进行解释。回归结果如下：

$$y = 8.296\ 60 + 0.227x_1 + 0.057\ 12x_2 - 0.000\ 679\ 55x_3 \tag{6}$$

式中的$y$是儿子的长期收入对数，$x_1$是父亲的长期收入对数，$x_2$是父亲的年龄，$x_3$是父亲年龄的平方。在回归中包括父亲的年龄的原因是纠正一个人的长期收入在50岁时可能比25岁时要高这样一个事实。相反，我们不需要对儿子的年龄进行纠正，因为研究中的每一个儿子都有相同的年龄。

我们可以用前面例子中的公式（4a）来估计代际收入弹性，它等于0.227。这一弹性意味着在控制了父亲的年龄效应之后，父亲长期收入每上升1%，就会导致儿子的长期收入上升0.227%。有关弹性问题的进一步阅读，请见伊斯特威尔等人的著作（Eatwell et al.，1987）。

——Osiris Jorge Parcero

（沈崇麟译校）

## 参考文献


Eatwell, J., Milgate, M., & Newman, P.(1987). *The new Palgrave: A dictionary of economics.* London: Macmillan.

Lindahl, L.(2002). *Do birth order and family size matter for intergenerational income mobility? Evidence from Sweden* (Working Paper No. 5). Stockholm: Swedish Institute for Social Research.

# 主位/客位之分(Emic/Etic Distinction)

1954 年,语言学家肯尼思·派克从音位(phonemic)和语音(phonetic)两个词中创造了主位(emic)和客位(etic)两个术语。派克用主位来指称对某一文化群体内部成员具有意义的内部文化特质,而客位指的是对研究者具有意义的外部观念和分类。

例如,现代医学用对任何病人都一样的简洁的、与文化无关(客位)的方式来定义“疾病”,而传统上人们会根据特殊的文化背景(主位地)对“疾病”作出不同的定义。西方植物学家根据林奈的分类法按客位给动植物分类,而当地人则根据他们独有的民间科学的世界观来对动植物进行主位的分类。有些巴布亚新几内亚(Papua New Guinea, PNG)的文化把蝙蝠归为“鸟”,而不是“哺乳动物”,因为蝙蝠飞起来像鸟一样。光学物理学家根据光的波长从客位来划分色谱,而本地人则以其地方的语言和文化为基础,按主位的意义区别对颜色进行不同的划分。比方说,英语中的黄、橙、棕,巴布亚新几内亚的 Selepet 人都用一个词来表示。

1964 年,人类学家马尔温·哈里斯从派克那里借来了这两个术语,将其用于他自己研究文化的文化唯物主义【Cultural Materialism】。他误解了派克,对这两个术语的使用与派克非常不同。派克曾经主张,如果要想正确理解某种文化,就要用主位方法来研究。哈里斯却反过来将主位方法定义成本地人为自己的风俗习惯所找的理想化的原因,而他用客位来指称风俗习惯在潜意识中的真正原因。例如,雅诺马马印第安人说他们攻击敌人的村庄是为了绑架年轻妇女(按哈里斯的定义,是一个主位的解释);而实际的客位原因,也即印第安人不承认的,就是敌人的村庄正在侵吞攻击方所需要的珍稀的野生猎物资源。

认知人类学家(cognitive anthropologist)在把主位/客位概念当作研究方法来使用时承袭的是派克的定义,而文化唯物论者则用了哈里斯改造过的方法。直到 1970 年代,人类学家和语言学家中对主位和客位究竟意味着什么都还是迷惑不清。到了 1980 年代,这两个术语出现在许多社会科学家的作品里。术语的含义也在更多的方向上大为增加,一部分是因为主位/客位的概念现在被用于其他的学科(包括心理学、社会学、医学、精神病学、经济学和宗教)。这些新的含义在黑德兰、派克、哈里斯(Headland, Pike, & Harris, 1990, pp.15-23)那里有所描述。

派克和哈里斯对主位和客位的定义中有一个共同点,就是他们都承认对一个文化的主观描述并不一定就是本地人意识到的模型。尽管许多定义都称主位模型是局内人的观点,也不总是这样。主位建构可能对持有这种观点的人有意义,而这些人经常意识不到,除非一个局外的民族志学者把他们自己的主位建构解释给他们听。

——Thomas N. Headland<br>Kenneth A. McElhanon<br>(杨可译 沈崇麟校)

*也可参见成分分析【Componential Analysis】。

## 参考文献

Harris, M. (1976). History and significance of the emic/etic distinction. *Annual Review of Anthropology*, *5*, 329-350.

Headland, T. N., Pike, K. L., & Harris, M. (1990). *Emics and etics: The insider/outsider debate.* Newbury Park, CA: Sage.

Jahoda, G. (1977). In pursuit of the emic-etic distinction: Can we ever capture it? In Y. H. Poortinga (Ed.), *Basic problems in cross-cultural*

*research* ( pp. 55-63 ). Amsterdam: Swets & Zeitlinger.

Lett, J. ( 1987 ). The importance of the emic/etic distinction. In *The human enterprise: A critical introduction to anthropological theory*. Boulder, CO: Westview.

Pike, K. L.( 1967). *Language in relation to a unified theory of the structure of human behavior*( 2nd ed.). The Hague: Mouton. ( Original work published 1954.)

## 情感研究(Emotions Research)

情感研究关注的是理解情感体验中的社会、心理和生理学过程,情感的社会建构(包括历史上的和跨文化的),社会生活中情感的作用,社会互动中情感的管理,以及情感在研究过程本身中的作用。

这个子领域的研究者在收集有关情感的资料时所使用的方法,与那些在社会学和社会心理学中的其他分支使用的方法非常类似,包括社会调查、实验、田野研究和深度访谈等。他们的分析中包含了各种定量和定性研究技术,其中一些研究路径值得特别注意。

在一些基于情感控制理论( affect control theory, ACT)( Heise, 1987)的研究中,对情感的界定是:那些从行动中产生的,对在进行互动中的人们肯定或否定的身份定位( confirmed or disconfirmed identities)。当社会行动者接收到基于否定的社会定位引发的情感暗示时,可能会通过重新定义情境或他人的身份定位来作出反馈。ACT 的资料收集通常包括要求参与者给三个维度的概念打分,即评价(好与坏)、效能(强劲与虚弱)、活力(活跃与怠惰)。——使用一种语义差别量表【Semantic Differential Scale】。绝大多数使用 ACT 的分析都是定量的。

在定性的、基于田野和访谈的情感研究中,研究者认为社会行动者的情感呈现是在情境中发生的,因而研究者会记录参与者对感知到的情感的描述,这样的情感是无法仅凭单独的观察来推断的。研究者同时也会记录那些从情感有关的互动中展露出来的结果。那些将会话分析【Conversation Analysis】技术应用于研究面对面沟通过程中的情感的研究者,可能会检查那些非常具体的交流。田野研究者和深度访谈研究者则与此不同,他们会发掘一系列的情境和条件,以便理解在与其他社会概念和过程相连接时情感是如何发挥作用的,这些社会概念和过程包括认知、感知、社会关系、社会地位、群体动力学或社会运动。

社会科学家过去在研究过程中追求的那种保持“情感中立”的做法,受到了越来越多的批评,现在许多田野研究者都将自己包括在研究参与者之内,而不是仅仅作为一个超然的、没有感情的观察者。他们把自己在田野工作中的情感体验作为非常有价值的资料来对待,并且在他们的田野笔记【Fieldnotes】和定性分析中,加入了具有反思性的观察和洞见( Kleinman & Copp, 1993)。那些将情感作为研究资料的研究者,可以获得更为深刻的社会学洞见,这些洞见有关他们的研究参与者和研究场景、参与者活动所受到的社会限制,以及他们的研究具有的社会意义等。

为了研究鲜活的情感体验和更好地理解人们是如何把情感作为复杂的过程那样来体验的,田野研究者可能会使用系统反省的社会学研究技术( Ellis, 1991)。这个

方法包括记录个体对于人、事物、事件的情感反应，然后分析这些感受的意义，以及它们是如何体现在持续发生的社会情景和文化实践的。系统反省方法可以由训练有素的参与者来进行，也可以由研究者作为一种自传式民族志【Autoethnography】的研究工具实施。

——Martha Copp

（郑晓娟译　沈崇麟校）

＊也可参见反身性【Reflexivity】。

## 参考文献


Ellis, C. S. (1991). Sociological introspection and emotional experience. *Symbolic Interaction*, *14*, 23-50.

Heise, D.R. (1987). Affect control theory: Concepts and model. *Journal of Mathematical Sociology*, *13*, 1-33.

Kleinman, S., & Copp, M.A. (1993). *Emotions and fieldwork*. Thousand Oaks, CA: Sage.


# 经验主义(Empiricism)

与实证主义【Positivism】类似，社会科学家现在使用“经验主义”这个术语的方式经常显示出对其较为负面的评价。的确，这两个词语经常被认为是同义词。这两个词现在都代表了这样一种观点，即由于科学所得出的结论是以合乎逻辑的方式，从经验资料中获得，因而科学才是获得知识的唯一真正途径，而这样的观点现在受到了质疑。

在认识论【Epistemology】中，“经验主义”代表了一种独特的研究方法，认为所有的知识都来自理智，与之相对，理性主义者【Rationalist】则认为存在天生的知识(Hamlyn,1967)。很大程度上，经验主义兴起于17世纪的科学革命，它强调对物质世界的经验性(特别是实验性)研究是获取科学知识的关键。这是对于早期研究过程的理念的背离，早期研究一直依赖古籍文本，宗教信仰和/或系统的哲学思维。

在这里可能有必要强调一下17世纪经验主义的激进和进步特征。这是由于当时的经验主义既包括了对之前有关知识的论断的批评，也包括了对于知识如何确证的方法的批评。此外，它还要求对这些批评的回应，必须秉持所有人均可获得证据的原则。这样，所有形式的权威都有可能受到潜在的挑战。

知识的经验主义理论所具有的意义和效度一直都在反复引起哲学的争论。对此持反对意见的人指出，如果抛开研究者或观察者的判断和解释，他们根本就无法获取任何研究资料，因而没有任何资料的效度能够被认定为是不容置疑的。同样，他们还认为科学不应该，也不能够简化为仅仅谋求从观察资料中符合逻辑地推断出法则的操作。相反，它需要对理论的发展，然后再针对数据进行检验；这是一个可能指明真伪，而非建立排除一切怀疑的效度【Validity】的过程。

在社会科学的语境下，经验主义大体上指的是一种在研究中过于强调经验资料，忽视理论建构的研究路径。一个例子是赖特·米尔斯对“抽象的经验主义”的批评(Mills,1959)。经验主义被认为与一种对证据的过分狭隘理解有关，指的是研究中把证据限制在了可观察或者可量化的

内容。这表明,“经验论”作为一个受到批评的术语,最常出现在定量研究【Qualitative Research】中,当然也有其应用于定性研究的例子。例如,一些民族志【Ethnography】研究,由于过分注重表面现象受到了批判理论学者的明显批评,他们记录了对世界的常识性解释,因而忽视了产生这些理想型表面现象的深层结构。另一方面,尽管很少有人试图为任何形式的狭义经验主义辩护,但我们却认为,今天对于经验主义的批评已经有点言过其实了,是在鼓励将推测的理论陈述作为事实的错误做法。

——Martyn Hammersley

(郑晓娟译　沈崇麟校)

### 参考文献


Bryant, C.G.A.(1985). *Positivism in social theory and research*.London: Macmillan.

Hamlyn, D. W. (1967). Empiricism. In P. Edwards (Ed.), *The encyclopedia of philosophy* (pp. 499-505).New York: Macmillan.

Mills, C.W.(1959).*The sociological imagination*. New York: Oxford University Press.


## 空格(Empty Cell)

在列联表【Contingency Table】的格【Cell】中没有个案时,就会发生空格。对空格造成的问题的解释,与含有数目非常少的格子造成的问题相仿。

——Alan Bryman

(沈崇麟译校)

* 也可参见稀疏表【Sparse Tables】。

## 编码/解码模型(Encoding/Decoding Model)

传媒研究中的编码/解码模型最初是伯明翰大学当代文化研究中心(the University of Birmingham's Centre for Contemporary Cultural Studies)的斯图尔特·霍尔(Hall, 1974)在1970年代构想出来的,在文化研究和大众传媒的社会学研究中很有影响。这一模型召回了沟通模型中最根本的三分结构,即发送者、信息、接收者,但突出了相对自主的编码和解码环节,并特别强调电视节目的文本多义性(多重意义)。最终,任何信息的意义都不是由信息传播的意愿决定的。而这不能被假定为受众只是简单地误解了原本想传递的含义。正如艾柯·翁贝托(Eco, 1965/1972)所提出的,大众传媒的“异常解码”是很正常的,因为现代媒体要面对的是数量如此巨大且社会背景又是如此不同的受众,他们不可避免地会将其不同的理解框架载入对信息的阐释。

对媒体信息的异常解码不只是个人差异,根据霍尔的说法,还有社会层面的原因。在资本主义社会中,阶级差异特别重要,通过比较价值体系和话语得以表现。霍尔使用了弗兰克·帕金(Parkin, 1973)对统治的、服从的和激进的意义/价值体系之间的区分来检验他称为统治的、协商的和对立的解码。我们有理由假设,大众传媒通常呈现

的是主导的——实际上是“中产阶级”的——价值观和理解世界的方式的合法性。因此，受到偏爱的意义和文本就会按常规被编码成媒体文本，通常它们都代表有权者的意识形态旨趣。而那些相对无权者，既可以接受已编过码的统治者的意识形态框架，也可以拒不接受，比如说既可接受，也可拒绝在电视新闻节目中的意识形态框架编码。例如，对于从管理的视角来报道的劳资冲突，属于左派的工人可能会将其解读成有偏见的节目。在直接选取优先的文本（统治的解码）还是从根本上质疑它（对立的解码）之间，受众还可以做其他各种各样的协商解码。

霍尔的编码/解码模型是文化和媒体研究中霸权理论范式的关键特征。这是对经典马克思主义主要的意识形态论点的重要修正。霍尔希望根据传播的渠道纯粹从语义上来探究意识形态过程的复杂性，同时在一定程度上保留社会中主导意识是统治阶级意识的论点。显然，社会中的从属群体不只是受统治意识形态的灌输，也可能会依据其自身的阐释框架、经验和旨趣来质疑其正确性。这样，意识形态力量就是在为社会领导权而斗争中逐事逐题的协商，而不只是来自上面的强迫命令。

编码/解码模型是一种启发手段，旨在把霸权理论用于批判式媒体研究。霍尔的一个学生大卫·莫利（Morley，1980）将其用于研究在教育环境中不同阶级背景的群体是如何解读1970年代英国电视上一个名为“全国”（*Nationwide*）的新闻杂志节目的。他得到了两个重要结论，对有关编码和解码日后的研究发展有着重要影响。首先，在差分解码中，阶级并不是唯一的，也不一定是最重要的因素。民族、性别和世代也是重要的社会学决定因素。在莫利及其追随者后来的工作中，对媒体类型的精细差分解码识别得到了蓬勃发展，以致首选意义、概念，及其暗含的意识形态批判都被削弱了。其次，莫利（Morley，1986）自己也开始对焦点小组【Focus Group】研究的不自然感到不满，尽管他自己也曾在对“全国”的研究中使用过这种办法。他对媒体使用的民族志【Ethnography】更感兴趣，尤其关心那些围绕着家里的电视的性别关系以及积极受众的想法。随着编码/解码模型的发展，它对大众媒体意识形态能量的质疑渐渐少了，而更多地去关注家庭内部的微观政治学和主权消费（sovereign consumption）的实际兴衰（完整的批判性评价参见 McGuigan，1992）。

——Jim McGuigan

（杨可译　沈崇麟校）

## 参考文献


Eco, U. (1972). Towards a semiotic inquiry into the television message. In *Working papers in cultural studies* (Vol. 3, pp. 103-121). Birmingham: Centre for Contemporary Cultural Studies. (Original work published 1965)

Hall, S. (1974). The television discourse—Encoding and decoding. In *Education and culture* (Vol.35, pp. 8-14). Paris: United Nations Educational, Social, and Cultural Organization.

McGuigan, J. (1992). *Cultural populism*. London: Routledge.

Morley, D. (1980). *The "Nationwide" audience*. London: British Film Institute.

Morley, D. (1986). *Family television—Cultural power and domestic leisure*. London: British Film Institute.

Parkin, F. (1973). *Class inequality and political order*. London: Paladin.

# 内生变量(Endogenous Variable)

一个内生变量是因果模型【Causal Model】或因果系统中的一个因子,它的值取决于该系统中其他变量的状况,它与其值取决于系统外部的外生变量不同。那些有关联但非等价的特质就是因变量【Dependent Variable】和自变量【Independent Variable】之间,以及待解释项和解说项之间的那些特质。一个因子,只是相对于模型设定【Specification of a Model】而言才可被归为内生的或外生的。该模型则代表在一组因果因子 $\mathbf{X}(x_1,x_2,\cdots,x_k)$ 中产生结果 $\mathbf{y}$ 的因果关系,其中 $\mathbf{y}=\mathbf{M}(\mathbf{X})$。如果变量 $x_j$ 的值取决于或受一个或多个自变量 $\mathbf{X}$(不包括它自身)的影响,那么我们就说这个变量在因果模型 $\mathbf{M}$ 中是内生的。一个纯粹的内生变量,是一个完全取决于系统内其他变量的状况的因子(如果一个因子是完全内生的,那么在理论上我们就可用代表 $x_j$ 组成的函数形式,作为 $\mathbf{X}$ 的一个函数来取代这一因子出现)。然而在实际的因果系统中,可能有着一系列内生变量。有些因子既受到系统内的因子的因果影响,又受到不包括在模型内的因子的因果影响。因此,一个给定的因子可能有一部分是内生的,而有一部分却是外生的——部分,而非全部取决于模型中其他变量的值。

我们来考量一个简单的因果系统——农耕系统。我们感兴趣的结果(解释因变量或待解释项)是玉米产量。许多因子(自变量、解说项)会影响玉米的产量。其中劳作、农夫的技能、可用的种子品种、可用的信用、气候、天气、土地的质量和类型、灌溉、害虫、气温、农药和肥料、畜牧业的饲养方法(animal practices)、可用的机动牵引力等因子尤为重要。这些变量都在一个特定的方面与玉米产量有着因果的关联。如果我们在一系列试验中改变这些变量水平,那么玉米的产量也将随之发生变化(增产或减产)。这些因子都对玉米产量有真正的影响,因而确定各种因子的性质和权重的确是一个真正的科学问题。而我们可能也注意到,在某些,但不是全部因子之间也有着因果关系。例如,虫害的水平受到降雨量、肥料(正面)和杀虫剂、劳作及技能(负面)的影响。因此,虫害部分是系统内内生的,部分是外生的,因为它也受那些系统外因子的影响[如平均气温、害虫载体的存在、(害虫)捕食者的下降]。

内生变量的概念尤其适用于时间序列的因果分析。在一个因果系统中,我们经常见到有些因子在时间 $n$ 的值取决于其他因子在时间 $n-1$ 时的值。假设虫害的水平独立于一给定时间的所有其他的因子,但却受到前一时间的降雨量和肥料的影响。在这种情况中,正确的说法是虫害在该时间内是外生的,但在历时上却是内生的。

——Daniel Little

(沈崇麟译校)

*也可参见外生变量【Exogenous Variable】。

## 参考文献

Hendry, D.F.(1995). *Dynamic econometrics.* Oxford, UK: Oxford University Press.

Pearl, J.(2000). *Causality: Models, reasoning, and inference.* Cambridge, UK: Cambridge University Press.

# 认识论(Epistemology)

哲学中的认识论指的是一种知识的理论,一种有关人类是如何逐渐获得关于其周边世界的知识——我们如何知道我们知道什么——的理论。认识论提供了一个哲学基础,在此之上我们才能确定何种知识是可能的,知识如何才能被判定是适当的和正统的。在社会科学中,这个术语用于确定究竟什么样的科学程序才能产生可靠的社会科学知识。

自从17世纪科学革命以来,哲学话语中有两种主导的认识论:理性主义【Rationalism】(以笛卡儿、莱布尼兹和斯宾诺萨为代表)和经验主义【Empiricism】(以约翰·洛克、乔治·贝克莱和大卫·休谟为代表)。其关注的问题就是为科学知识寻找可靠的基础,以将其与信仰和偏见相区别。理性主义基于这样一个观念:可靠的知识都先源于"纯粹"的推理、无争论余地的公理,然后再用形式逻辑得出结论。从这种观点来看,数学产生的是这样一种知识。而另一方面,经验主义则依赖人类的感觉来生产可靠的知识。这意味着只要借助直接的感官经验就可以获取有关世界的知识。

在社会科学的背景中,这些哲学立场可以用两种主导的认识论立场及其相关的本体论立场(参见本体论【Ontology】)来进一步详细阐释。按第一种认识论立场,也即唯名论来看,用于描述和解释的概念只是被当作便利的集合名词,它们被发明出来作为对被观察事物的一般范畴的概括,例如,"社会行动者"或"社会群体"。这些集合名词不应与现实本身混淆,也不具有行动的能力;现实是事件组成的,只有个人才可能行动。按第二种认识论,也即唯实论【Realism】来看,科学概念就是用于不一定能被观察到的揭示有关社会现实的某些未必是可观察的事物的。这类概念就是设计出来透过可观察的事件,看到构成和解释那些事件的真实所在(也请参见批判实在论【Critical Realism】)。

当唯名论和唯实论及两种主要的可供选择的本体论立场,也即唯物论和观念论结合在一起时,就产生了一个四项的分类方案。经验主义把唯物的本体论(参见唯心论【Idealism】)与唯名论的认识论结合在一起,实体主义(Substantialism)把唯物的本体论和唯实论的认识论结合在一起,主观主义把观念论的本体论和唯名论的认识论结合在一起,而理性主义则把观念论的本体论和唯实论的认识论结合在一起(Johnson, Dandeker, & Ashworth, 1984)。

经验主义把现实视为由可以用人类的感官来观察的物质事物组成的。概念和概括则是依据许多观察所做的简短总结。实体主义也采用了唯物论的现实观,但承认人们在不同的时间和地点可能会对现实作出不同的解读。然而,他们却把物质世界视为对人类行为和社会关系的束缚。因为主观主义拒绝物质世界的观点,并把现实看成是在社会建构和解读的,这种现实的知识只能从社会行动者给予它的叙述中获得。最后,理性主义把现实看成是真实的和一般的;它不依赖人、人的意识和环境而独立存在。因为这种现实是由观念组成的,获取有关知识的方法只能是通过考察思考过程、人类共享的天赋观念——简言之,即理智结构本身。这4种立场必须被视为理想类型,其间存在固有的张力(Johnson et al., 1984)。它们主要与社会科学哲学【Philosophies of Social Science】存在关联。经验主义与实证主义【Positivism】和

证伪主义【Falsificationism】相关，实体主义与批判实在论【Critical Realism】相关，主观主义则关系到诠释主义【Interpretivism】。在涂尔干有关自杀的研究中可以看到理性主义，但在今天的社会科学中已不太常见了。

——Norman Blaikie

（杨可译　沈崇麟校）

## 参考文献


Blaikie, N. (1993). *Approaches to social enquiry*. Cambridge, UK: Polity.

Crotty, M. (1998). *The foundations of social research*. London: Sage.

Johnson, T., Dandeker, C., & Ashworth, C. (1984). *The structure of social theory*. London: Macmillan.


# EQS

EQS 是用来做结构方程模型【Structural Equation Models】分析的软件。它以易于处理非正态和非连续的数据著称。有关这一软件的进一步信息请参见该软件的网站。

——Tim Futing Liao

（沈崇麟译校）

# 误差（Error）

在日常语汇中，“误差”的意思是“错误”。在数学语汇中“误差”的意思是计算中的不精确程度。在统计学中，“误差”一词可能并没有负面的含义。它通常代表实体的一种观察值（如样本均值），和它的“真”或“期望”值（如总体均值）之间的差异或矛盾。在很多情况下，这种差异均系偶然因素所致。正因如此，它常被称为“猜测的（stochastic）”误差或随机误差项。“stochastic”一词来自希腊语的 stokhos，意思是“公牛的眼睛”。往标靶上投掷飞镖的结果就是一个机遇过程，也就是说，这是一个担心错失的过程。在统计学中，这个词意味着存在着一个随机变量，即一个其结果取决于一个可能的实验的变量。

我们在实际工作中可能遇到若干种类型的误差。这些误差包括白噪声【White Noise】、第一类误差【Type I Error】、第二类误差【Type II Error】、测量【Measurement】误差、变量中的误差、方程中的误差、模型设定误差、丢失观察造成的误差和预测【Forecasting】误差。

## 白噪声误差

令 $u_t$ 为一随机误差项。如果它有零均值，常方差和无自相关（或序列相关【Serial Correlation】），那么它就称为白噪声误差项，用 $u_t \sim WN(0,\sigma)^2$ 表示。如果在序列相关之外，$u_t$ 还是序列独立的，那么 $u_t$ 就是独立白噪声，用 $u_t \sim iid(0,\sigma)^2$ 表示。表达式中 *iid* 的意思是独立和恒等分布的。如果除此之外，$u_t$ 还是正态分布的，那么它便称为正态白分布或高斯白噪声，用 $u_t \sim iidN(0,\sigma)^2$ 表示。也许还应该注意的是，只有在正态分布中，零相关才意味着独立。

白噪声在时间序列建模【Time Series

Modeling】中(参见预测)及经典线性回归模型中起着关键性的作用。

## 观察或测量误差

在收集数据时,因为错误的测量工具或人为的因素也会产生误差。一种常见的误差是舍入误差(rounding error)或近似误差(error of approximation)。在这种形式的误差中,不是对观察作记录,比如说,记录一个4位的小数点,而是将它四舍五入为3位。在调查型数据中,提问者可能因为回答人不回答问题,或有的问卷没有返回而得不到所有问题的信息。有时,变量的信息,如收入和财产这样的变量的信息往往比较难以得到,因为回答人害怕财税部门追究而不愿提供真实的信息。

## 变量中的误差

有时变量的信息的确是很难得到的。例如,期望的或预料的通货膨胀的数据之所以难以得到,是因为某人的通货膨胀的期望与他人的不一致。经济学家设计了各种方案,或代理变量来测量预料或期望的通货膨胀。有一种期望的通货膨胀的量度是通过调查主要的经济学家,了解他们对不久的将来的通货膨胀问题的看法得到的。这些经济学家的观点可能以数学模型为根据,有些则纯粹是个人的看法。再举一个经济学方面的例子。大家知道,"失业的自然增长率"这一变量是很难测量的。有些经济学家采用某种前几个月或季度的实际失业率的加权平均,把它作为失业的自然增长率的代理。所有这些代理的量度都有可能存在误差。

## 在检验假设中的误差

在检验统计假设时,有可能发生第一类误差或第二类误差。第一类误差是拒绝了一个真的假设这样的误差,而第二类误差则是未能拒绝一个假的假设这样的误差。这两类误差对 Neyman-Pearson,或经典假设检验理论而言都是最基本的。最理想的是能同时缩小这两类误差。然而,对一给定的样本量而言,我们是不可能同时缩小这两种类型误差的。经典的假设检验通常都是先在某个主观选择的显著水平【Level of Significance】上,如1%,5%或10%,设定第一类误差的概率,而不过多地考虑第二类误差。从这个意义上讲,判一个无罪的人有罪(的错误)远比漏判一个有罪的人严重。

## 方程误差

方程误差一般都与回归【Regression】分析有关。回归分析的目的是依据一个或多个$X$(解释变量【Explanatory Variable】)来解释变量$Y$(因变量【Dependent Variable】)的性状。无论考量的变量$X$有多少,我们都不可能以选取的$X$为依据对$Y$的性状作出完全的解释。$Y$的真值和那些用回归方程估计得到的值之间的差异代表方程的误差项。我们可用数学符号把回归方程写作

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \cdots + \beta_k X_{ki} + u_i$$

该方程阐明,无论模型中考量的解释变量($X$)有多少,在$Y$的变差中也必定有一些未能为所有的$X$变量所解释。在$Y$中必定有某些固有的随机性未能被解释,而这内在的随机性则由随机的【Stochastic】误差项$u_i$表示。这种误差项的产生可能有几种原因,如构成回归模型的理论的模糊性,某些解释变量的数据无法得到,或使用质量欠佳的代理(如使用实际的通货膨胀率代替失业模型中的期望通货膨胀率)。有时,即使研究者可能知道模型中所有的解释变量,但是为了遵循简约的原则(或奥卡姆剃刀原理),也

还是会决定只将少数几个核心的变量包括在分析之中。

随机误差项 $u_i$ 在回归分析中起着极为重要的作用,因为估计的回归系数——$\beta$ 系数的性质在很大程度上取决于有关随机误差项的假设。在经典回归模型中,我们假定 $X$ 或解释变量的值在重复抽样中是固定的,而误差项 $u_i$ 假定是高斯白噪声的。正因如此,估计的 $\beta$ 系数也是服从正态分布的。这使得检验真或总体 $\beta$ 系数变得更为简单。不仅如此,在关于误差项 $u$ 的正态性假定下,$\beta$ 系数的最大似然估计【Maximum Likelihood Estimation, MLE】和普通最小二乘【Ordinary Least Squares, OLS】估计量是相同的。

应该注意的是,在实际研究中,我们是观察不到误差项 $u_i$ 的,我们能观察到的反而是残差项 $\hat{u}_i$,它是实际的 $Y_i$ 值和它用回归方程估计的值之间的差异。换言之,$\hat{u}_i$ 是真 $u_i$ 的代理。我们可以证明,随着样本量无限地增长,残差将逐渐收敛于它们的真值。

## 模型设定误差

一种理论虽然可能给我们提出哪些变量可以包含在模型中的建议,但却常常无法提出那些变量应该进入的模型的特定函数形式。例如,我们来考虑一个这样的问题:某种产品的需求($Y$)与它的价格($P$)和消费者的收入($I$)有关。现在我们来考虑两种可能的(回归)模型

$$Y_i = \beta_1 + \beta_2 P_i + \beta_3 I_i + u_i,$$
$$\ln Y_i = \alpha_1 + \alpha_2 \ln P_i + \alpha_3 \ln I_i + u_i$$

式中的 $u_i$ 是随机误差项,而 ln 则是自然对数。

第一个模型是线性回归模型,而第二个模型则是对数线性回归模型。尽管两个模型含有相同的变量,但是它们输入模型的方式却造成了在模型解释上的很大差异。例如,在线性模型中的 $\beta_2$(偏斜率系数)是由 $P$ 值的变化而引起的 $Y$ 的均值的变化率,而对数线性模型中,$\alpha_2$ 测量的是 $Y$ 对 $P$ 的弹性(elasticity),也即它测量的是 $P$ 的(小的)百分比变化引起的 $Y$ 的百分比的变化。显然,这两个量度是不同的。

## 丢失观察造成的误差

研究者经常遇到丢失数据【Missing Data】或丢失观察的问题。这种问题在调查型数据中是相当普遍的,且一般都发生在回答人未回答所有的调查问题时。在时间系列数据中,我们可能会遇到几种方式的数据丢失。例如,某些时间系列的数据可得到分月的数据(如失业率),但另一些系列的数据只能得到分季度的数据(如国内生产总值或 GDP 的数据)。有时数据中因特殊况而存在一些缺口。例如,在第二次世界大战期间,有些数据就无法收集。

如果我们忽略那些丢失的观察,只对那些能得到的数据作分析,那么我们就可能得到感兴趣的参数的有偏倚的或前后矛盾的估计值。例如,我们不妨来看一下下面的简单的回归分析模型

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$

式中的 $Y$ 是闲暇旅游的支出,$X$ 是个人收入,而 $u$ 则是随机误差项。假设数据来自一个调查的 1 000 个个人。该调查问及回答人过去 6 个月中的旅游支出问题。我们再进一步假设,100 个个人没有报告任何旅行开支。如果我们决定只以 900 个个人的数据为依据作回归分析(一种与使用截删数据【Censored Data】类似的方法),那么我们可以证明 $\beta$ 系数的 OLS 估计量不仅是有偏倚的,而且是前后矛盾的。如果这 100 个个人的收入也未曾得到(这种情况

叫作删节,参见截删和删节【Censoring and Truncation】),那么情况会变得更加复杂。

### 预测误差

涉及时间序列【Time Series】的回归模型的目的之一是以给定的解释变量之值,用估计的模型预测因变量未来的值。

假设有人开发了一个用过去40个季度的数据解释美国汽车销售情况的回归模型。假定用通常的统计学标准来判断,模型是一个好模型。假设,我们想用这个模型预测未来5个季度的汽车销售情况。在实际销售的信息得到之后,预测的未来5个季度的汽车销售数据,它们可以和实际的销售情况进行比较。无论这个模型怎么好,都无法保证汽车销售的预测值和实际值将会是完全相同的,而它们之间的差异就是预测误差。

统计学家们已经开发出了各种各样的预测方法,所有这些方法都力图尽可能地降低预测误差。这一系列模型涵盖了从简单的移动平均数【Moving Average】模型到非常复杂的模型,如ARIMA[自回归求和移动平均(Autoregressive Integrated Moving Average)]和向量自回归【Vector Autoregression,VAR】等各种模型。在开发这些模型时,白噪声误差项起着关键的作用。为了评价各种预测模型的性能,研究者常常使用池田信息准则(Akaike Information Criterion,AIC)或西沃茨信息准则(Schwarz Information Criterion,SIC)这些准则的值越低,模型就越好。

——Damodar N. Gujarati

(沈崇麟译校)

### 参考文献


Darnell, A. C.(1995). *A dictionary of econometrics.* Cheltenham, UK: Edward Elgar.

Diebold, F. X.(2001). *Elements of forecasting* (2nd ed.). Cincinnati, OH: South-Western.

Gujarati, D.(2002). *Basic econometrics* (4th ed.). New York: McGraw-Hill.


## 误差纠正模型(Error Correction Models)

误差纠正模型(ECM)是一种时间序列【Time Series】回归模型,它以性状假定,即两个或多个时间系列展示了确定短期和长期性状的均衡关系的假定为根据。ECM最初由詹姆斯·戴维森、大卫·亨德里、弗兰克·斯尔芭、斯蒂芬·约(Davidson, Hendry, Srba, & Yeo)于1978年在经济学中普及。1987年,罗伯特·恩格尔和克里弗·格兰杰证明,共合时间序列(cointegrated time series)可很好地为一个误差纠正模型所表示。从那时起,ECM便开始与共合时间序列相关联。然而重要的问题在于要注意,ECM是一个自回归分布滞后(Autoregressive Distributed Lag, ADL)模型的一般转换。尤其是,ADL模型可以重写为一个ECM,这样在ECM表示法中便不存在对ADL参数的限定了。这样ECM便可用来为涉及平稳时间序列以及共合时间序列的平衡关系建模。

ECM模型由下式给出:

$$\Delta y_t = \lambda_0 + \lambda_1 \Delta x_t - \gamma(y_{t-1} - \beta x_{t-1}) + \varepsilon_t$$

式中的$\Delta y_t = y_t - y_{t-1}$,$\Delta x_t = x_t - x_{t-1}$,$\gamma$是误差纠正率,它给出了不平衡——($y_{t-1} - \beta x_{t-1}$)的被纠正的比率。括号中的那些项也可以包括

另外的自变量(Banerjee,Dolado,Galbraith, & Hendry,1993 有关各种 ECM 的一般讨论)。

假设我们认为消费者信心(consumers' sentiment)和公众对总统的经济管理能力的看法有联系:在人们比较确信总统的管理能力时,他们对经济也更加确信。换句话说,我们认为在经济赞许和消费者信心之间有着一种长期的平衡。如果新闻报道越来越多地批评总统的经济政策,致使公众重新评价他们对总统管理能力的看法(向下),那么目前的消费者信心可能就会过高。在这样的情形中,我们也许可以说,两个系列已经失去平衡,因而我们期望公众的经济信心降下来。我们可以在一个 ADL 模型或 ECM 模型中捕捉到这种关系。然而,如果我们希望这样的校正要多长时间——经济信心的反应是多快或多慢——我们只能用 ECM 来直接估计这种效应。经上面那样的参数化,误差纠正率由 $\gamma$ 给出。$\gamma$ 为 0.2 表明在下一个时期将有 20%的不平衡被纠正。在下下一个时期又有一个 20%被纠正,直至平衡恢复为止。一个 0.8 的 $\gamma$ 意味着更快的修正。因此,ECM 对误差纠正率的估计值直接进行了修正。此外,因为 ECM 在模型右侧包括了长期(水平)和短期(变化)变量,所以我们能同时捕捉到消费者信心对评估中的短期变化和长期水平的反应。ECM 能说明时间系列有多高或多低(有多么肯定或否定的评价)和移动的方向的特性正是它特别吸引人的地方。

在实际研究中,许多平衡模型是以永久记忆为前提的,因而不能用分布滞后模型估计。在这种情况下,ECM 一般都是首选的模型。在上述单方程表达式之外,分析常常会依靠用于共合情况的 ECM 估计的 Engle-Granger 的两步法(Engle & Granger,1987)。在这种情况下,分析者在第一步回归中估计平衡,然后再将第一步模型的残差(不平衡)输入第二步的误差纠正模型。

——Suzanna De Boef

(沈崇麟译校)

## 参考文献


Banerjee, A., Dolado, J. J., Galbraith, J., & Hendry, D. F. (1993). *Cointegration, error correction and the econometric analysis of nonstationary series*. New York: Oxford University Press.

Davidson, J., Hendry, D. F., Srba, F., & Yeo, S. (1978). Econometric modelling of the aggregate time-series relationship between consumers' expenditure and income in the United Kingdom. *Economic Journal*, *88*, 661-692.

Engle, R., & Granger, C. W. J. (1987). Co-integration and error-correction: Representation, estimation, and testing. *Econometrica*, *55*, 251-276.


## 本质主义(Essentialism)

在哲学和社会科学中,本质主义这个术语用于指称他人所谓的错误信念,即称自己已经准确地描述了某事物或某人的本质特征,或认为起码这应该是哲学或科学之目标的错误信念。将事物本质特征与其外在表现之间的传统哲学区分为现代社会科学的起源提供了材料,比如说,马克思就强调,如果事物真的就是其外在看似的那样,科学就没有必要了,而在社会研究中,“抽象的力量”代替了那些诸如化学分析这样的程序。

尽管有些评论家因马克思的本质主义而称赞他，称其可与亚里士多德的哲学媲美，但大多数评论家都把它看成是一种错误，无论他们是否认为马克思犯过这一错误。那些持有后一种观点的人，倾向于将他的观点称为唯心论【Realism】。

近年来，对本质主义的讨论基本上沿袭的是卡尔·波普尔（Popper，1966）的批评，这一批评首先在《开放社会及其敌人》（*The Open Society and Its Enemies*）中提出，波普尔在其中将方法论的本质主义称为“一种观念……，认为纯粹知识或‘科学’的任务就是发现和描述事物真实的本性，即它们隐藏的现实或本质”（Popper，1945/1966，p.31）。对这种观念，波普尔以其“可误论者”的立场加以反驳，他认为科学就是通过驳斥错误，依靠“逼真”而非真实才得到了更好的理解。

波普尔持有的则是另一种版本的唯实论，大多数的现代唯实论者认为他的唯实论不仅太弱，而且对本质主义的批评也是错误的。其他的唯实主义者，例如，罗姆·哈雷（Harré，1986，pp.103-105）和罗伊·巴斯卡尔（Bhaskar，1978）就曾借助自然的种类和自然的必然性等概念，且为在本质上对辩驳和进一步的发展开放的科学模型中理解的本质作过适度的辩护。其他的唯实论者，例如塞耶（Sayer，1997），也曾对唯实论遭受的本质主义抨击提出辩护，也主张各种适度的本质主义。对“科学性本质主义（dispositional essentialism）”（Sayer，1997）的技术哲学讨论也在社会科学中得到了回应，这种技术哲学观点声称，诸如带正电荷这样一个性质，在本质上关联着吸引某种带有负电荷的事物的倾向。塞耶反对基于可能仅仅是偶然的规律就作出“联想思考”，为唯实论的抽象提出了辩护（Sayer，2000，2001；Holmwood，2001）。

近年来，女权主义者和反种族主义者对本质主义（“所有女性/黑人在本质上都是xyz”）的批评有时会吸收解构主义者或“后结构主义者”的立场，而这二者的立场都是在形而上学和现象学对本质主义的批判中发展起来的。不过，其他的女权主义者，例如弗斯（Fuss，1989）对本质主义和社会建构论之间的二分法提出了批评。因此，大量这类争论都和方法论争论有交叉；比如说，德兰蒂（Delanty，1997）曾主张，要在社会科学的哲学中把唯实论和建构论或相对论的立场综合起来。

——R. William Outhwaite

（杨可译　沈崇麟校）

## 参考文献


Bhaskar, R. (1978). *A realist theory of science.* Hassocks, UK: Harvester Press.

Delanty, G. (1997). *Social science: Beyond constructivism and realism.* Buckingham, UK: Open University Press.

Ellis, B. D., & Lierse, C. (1994). Dispositional essentialism. *Australasian Journal of Philosophy*, *72* (1), 27-45.

Fuss, D. (1989). *Essentially speaking.* New York: Routledge.

Harré, R. (1986). *Varieties of realism.* Oxford, UK: Blackwell.

Holmwood, J. (2001). Gender and critical realism: A critique of Sayer. *Sociology*, *35*, 947-965.

Popper, K. (1966). *The open society and its enemies* (5th ed., Vol. 1). London: Routledge. (Original work published 1945.)

Sayer, A. (1997). Essentialism, social constructionism and beyond. *Sociological Review*, *45*, 453-487.

Sayer, A. (2000). System, lifeworld and gender: Associational versus counterfactual thinking. *Sociology*, *34*, 707-725.

Sayer, A. (2001). Reply to Holmwood. *Sociology*, *35*, 967-984.

# 估计(Estimation)

统计的一个主要目的是用样本【Sample】的信息对总体【Population】作推论。假设我们的兴趣在于研究一个有100 000居民的美国小城的成年人口的年收入。那么总体将由所有在研究时有经济活动的成年人组成。城中成年人口的年收入可用被称为参数【Parameter】的各种数字量度来描述。参数的例子包括成年人口的年收入均值【Mean】、收入的方差【Variance】(一种变异性量度)和个人的教育与其收入之间的关联【Association】。统计分析的一般目的是对一或多个总体参数进行推论。本文将要讨论的是总体参数的估计问题。在对点和区间估计进行了简短的讨论之后,我们将进行普通最小二乘估计量【Estimator】和最大似然估计量的讨论。通过讨论,我们将用一个简单的数字例子来阐述这些估计量背后的基本思想。几乎所有的教科书都在这方面提供了更为一般的处理方法。

## 点估计和区间估计

城中成年人口年收入统计调查可给出两种形式的估计量。例如,可给出样本均值为45 500美元。样本均值 $\bar{Y} = \left[\sum_{i=1}^{n}(Y_i)\right]/n$ 是一个单个的数字或点估计值。我们希望这个估计值与总体均值 $\mu$ 相去不远。作为另外一种选择,调查也可以给出两个数字的估计值。例如,(42 000美元,49 000美元),意思是城中人口的平均年收入可能在42 000美元和49 000美元之间。这种备择的估计形式被称为区间估计,被设计用来围住正在研究的参数。一般区间估计量包含总体参数的概率要高于点估计量。点和区间估计都必须由一个估计量产生。而估计量则是用一个样本来计算估计值的程序或方法。

## 普通最小二乘估计量

普通最小二乘【Ordinary Least Squares, OLS】估计量常被用于估计线性回归中的参数。这种回归关注的问题是一个连续的因变量 $Y_i$ 是怎样与自变量 $x_{1i}, x_{2i}, \cdots, x_{ki}$ 有关的。诸如这样的估计量可取如下形式:

$$Y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki} + e_i \quad (1)$$

在一个线性回归只有一个自变量时,它是像下面这种形式的简单的线性回归

$$Y_i = \beta_0 + \beta_1 x_i + e_i \quad (2)$$

这样一种回归可以用来研究总体中的年收入是如何与教育年限有关的。图1是一张假设的统计图,它是一张用来自城市总体的样本数据绘制的对照教育年限的年收入散点图。平均而言,那些有着更多教育年限的人倾向于有较高的收入。虽然有许多数学函数可以被用来给年收入和教育之间的关联建模,但统计实践通常都是用线性模型,譬如像方程(2)这样的线性模型。在这一特定的例子中,$\beta_0$ 就是那些没有受过任何正规教育的人的年收入量,而 $\beta_1$ 则是和增加一年教育年限关联的年收入量。这样方程(2)就描述了总体中年收入和教育之间的关系。所谓统计调查就是用总体的样本来计算 $\hat{\beta}_0$ 和 $\hat{\beta}_1$,它们是总体参数 $\beta_0$ 和 $\beta_1$ 的估计值。

用样本来估计 $\hat{\beta}_0$ 和 $\hat{\beta}_1$，实际上是为了拟合一条通过图 1 中那组数据点的直线。我们知道直线是无法预测所有的数据点的，因为教育并未完全对收入作出预测。我们希望这条直线总体而言，其所处的位置能最好地代表这组数据以这样一种方式来放置。我们可以用眼睛来拟合这条直线，但这只能近似地拟合简单的线性回归或双变量问题。更为系统的拟合方法是借助 OLS 估计量，来产生如下的拟合直线。

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i \qquad (3)$$

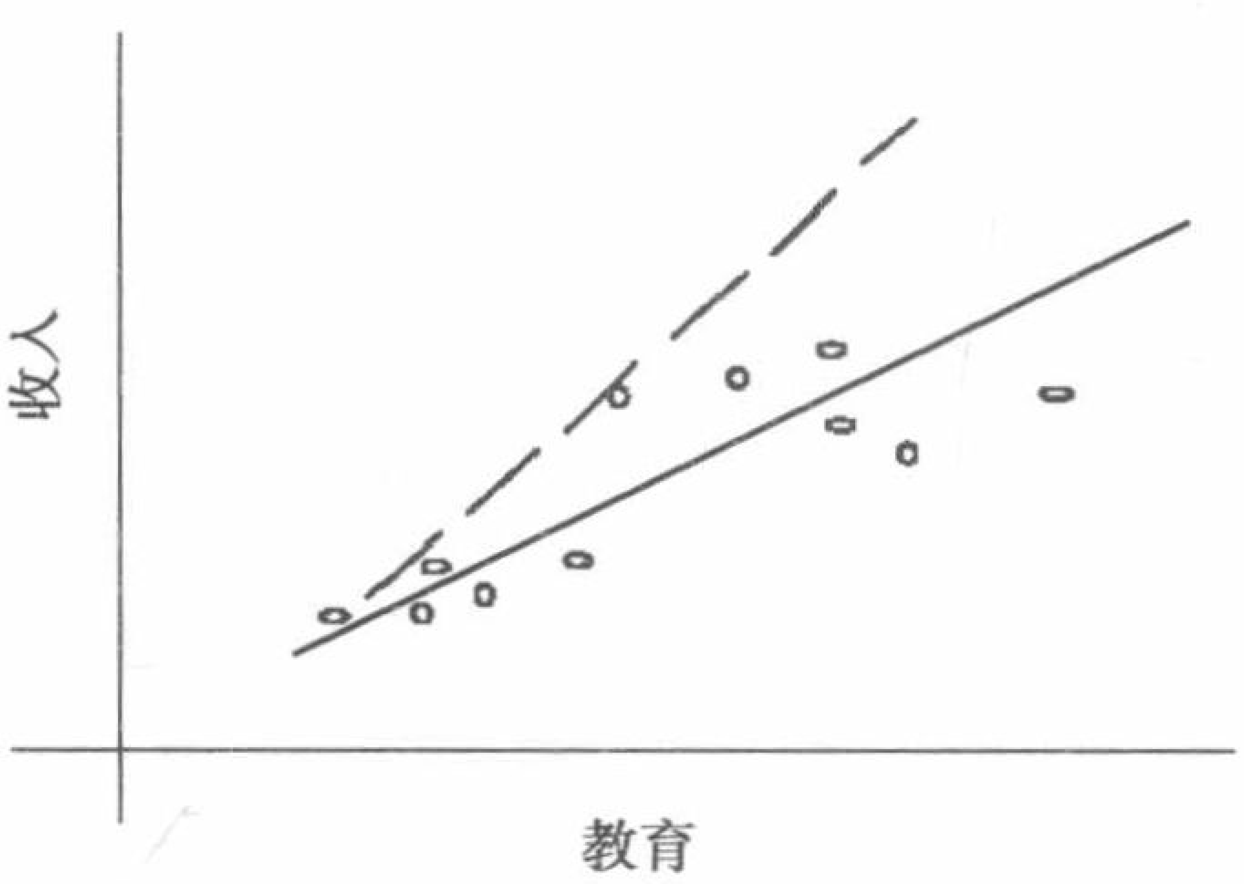



图 1　收入对教育的散点图和拟合直线

观察的数据点 $y_i$ 和与由 $\hat{y}_i$ 表示的拟合直线之间的垂直距离（或 $y_i - \hat{y}_i$）被称为偏差或预测误差。OLS 估计量的基本思想是选择那个使我们所有偏差之和 $\sum_{i=1}^{n}(y_i - \hat{y}_i)$ 最小的 $\hat{\beta}_0$ 和 $\hat{\beta}_1$。但是有些数据点位于直线之上，有些却位于直线之下。位于直线之上的正误差常常会取消位于直线之下的负误差，余下的所有偏差的和会远远小于它本来应有的值。解决这一问题的规范的方法是先取每个误差的平方，再极小化误差平方和（SSE）。

$$\text{SSE} = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{n}[y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)]^2$$

对应于 SSE 的最小值的 $\hat{\beta}_0$ 和 $\hat{\beta}_1$ 可以通过解方程 $\partial\,\text{SSE}/\partial\hat{\beta}_0 = 0$ 和 $\partial\,\text{SSE}/\partial\hat{\beta}_1 = 0$ 求得。取关于 $\hat{\beta}_0$ 和 $\hat{\beta}_1$ 的偏导数，并将结果置为 0，我们得到：

$$\frac{\partial\,\text{SSE}}{\partial\hat{\beta}_0} = -2\left(\sum_{i=1}^{n} y_i - n\hat{\beta}_0 - \hat{\beta}_1\sum_{i=1}^{n} x_i\right) = 0$$

和

$$\frac{\partial\,\text{SSE}}{\partial\hat{\beta}_1} = -2\left(\sum_{i=1}^{n} x_i y_i - \hat{\beta}_0\sum_{i=1}^{n} x_i - \hat{\beta}_1\sum_{i=1}^{n} x_i^2\right) = 0$$

这两个方程是 OLS 方程，可同时对 $\hat{\beta}_0$ 和 $\hat{\beta}_1$ 求解：

$$\hat{\beta}_1 = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x}$$

假设我们的样本由 8 个个体组成，他们的年收入分别为 35，57，26，45，58，40，55 和 48（以千美元计），而教育年限分别为 12，18，11，12，14，12，16 和 15。表 1 阐述了参数 $\hat{\beta}_0$ 和 $\hat{\beta}_1$ 的 OLS 计算方法：

$$\hat{y}_i = -7.18 + 3.83x_i \qquad (4)$$

表 1　最小二乘拟合的直线的计算

| 个体 $i$ | $y_i$ | $x_i$ | $x_i y_i$ | $x_i^2$ |
|---|---|---|---|---|
| 1 | 35 | 12 | 420 | 144 |
| 2 | 57 | 18 | 1 026 | 324 |
| 3 | 26 | 11 | 286 | 121 |
| 4 | 45 | 12 | 540 | 144 |
| 5 | 58 | 14 | 812 | 196 |
| 6 | 40 | 12 | 480 | 144 |

续表

| 个体 $i$ | $y_i$ | $x_i$ | $x_i y_i$ | $x_i^2$ |
|---|---|---|---|---|
| 7 | 55 | 16 | 880 | 256 |
| 8 | 48 | 15 | 720 | 225 |
| 总计 | $\sum y_i = 364$ $\sum x_i = 110$ $\sum x_i y_i = 5\,164$ $\sum x_i^2 = 1\,554$ | | | |

因此,斜率的 OLS 估计值显示,教育每增加一年,将会多增加 3 830 美元的年收入。

## 最大似然估计量

最大似然估计量(maximum likelihood estimator, MLE)是大多数用于社会科学的统计模型,如 Logistic 回归、Probit 回归、多项回归、Poisson 回归、负二项回归、各种事件史模型、成长曲线模型和多水平线性回归使用的统计量。在过去的 15 年间,期望最大化(EM)逐渐普及,基本上已经成为最大似然估计的一种计算方法。甚至前面讨论的 OLS 的参数估计,在数据具有正态分布的假设下,也可被看成是最大似然估计。假设我们希望知道小城中成年人口的平均年收入。估计这个值的一个可行的方法是抽取和应用一个随机样本。假设我们的样本由一个有如早先设定的那样的收入的 8 个个体组成(最大似然估计实际上应该用于大样本,在这里使用 8 个个体的小样本只是为了简化问题的阐述)。为了继续进行最大似然估计,我们必须设定数据的分布【Distribution】。到目前为止,大多数线性数据假设的分布都是正态分布。因此,我们假设 8 个个体的收入的每一项都来自一个正态分布【Normal Distribution】:$Y_1 \sim N(\mu, \sigma^2), \cdots, Y_8 \sim N(\mu, \sigma^2)$,有密度

$$f(y_i) = \frac{1}{\sqrt{2\pi\sigma^2}\, e^{-\frac{(y_i-\mu)^2}{2\sigma^2}}}$$

参数 $\mu$ 并非只是简单的平均年收入。相反,它是表示总体年收入正态分布的均值。

最大似然估计的下一步是写出数据的似然函数。个体 $i$ 的似然函数是被看作未知参数的函数的个体的密度

$$L_i(\mu; y_i) = \frac{1}{\sqrt{2\pi\sigma^2}\, e^{-\frac{(y_i-\mu)^2}{2\sigma^2}}}$$

假定样本中的个别观察是独立的,那么我们可以将整个样本的联合似然函数写作:

$$L(\mu; y_1, y_2, \cdots, y_8) = \prod_{i=1}^{8} L_i$$

$$= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^8 e^{-\sum_{i=1}^{8}\frac{(y_i-\mu)^2}{2\sigma^2}}$$

MLE 是将联合似然函数 $L(\mu, y_1, y_2, \cdots, y_8)$ 最大化的 $\mu$ 的值。而实际上,将 $\ln L(\mu)$,$L(\mu)$ 的自然对数最大化,要比将 $L(\mu)$ 最大化容易得多。与似然函数 $L(\mu)$ 关联的数量,对计算而言,常常会变得太小或太大。我们之所以有理由这样做,是因为任何的对数似然最大化,也必定会将似然最大化。采用对数之后,$L(\mu)$ 被简化为:

$$\ln L(\mu) = 8 \ln \frac{1}{\sqrt{2\pi\sigma^2}} - \frac{\sum_{i=1}^{8}(y_i - \mu)^2}{2\sigma^2}$$

求 MLE 的值要取关于 $\mu$ 的对似然函数的偏导数,将导数置为零,然后再求解未知参数的方程:

$$\frac{\partial \ln L(\mu)}{\partial \mu} = \frac{\sum_{i=1}^{8} 2(y_i - \mu)}{2\sigma^2} = 0$$

$$\hat{\mu} = \sum_{i=1}^{8} \frac{y_i}{8} = \bar{y}$$

$$= \frac{35 + 57 + 26 + 45 + 58 + 40 + 55 + 48}{8}$$

$$= 45.5$$

以上是得到的结果，我们也可以用类似的方法求得 $\sigma^2$ 的 MLE：

$$\frac{\partial \ln L(\mu)}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}\sum_{i=1}^{8}(y_i - \mu)^2 = 0$$

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{8}(y_i - \bar{y})^2$$

因此，正态分布的均值和方差的 MLE 便成了相应的样本的特征。在简单模型中，MLE 常常是"明显的"估计量，而在模型变得稍微复杂一点之后，这种明显性一般就会丢失。因我们可能想把最简单情形中的 MLE 看成是简单和易于理解的，所以最简单情形中 MLE 的明显性也许可以看成是可靠的。

最大似然估计是一种有直觉吸引力的方法。且不说在最简单情形中的 MLE 的明显性，MLE 是最有可能产生观察样本的总体估计量的。更为重要的是，MLE 之所以如此受欢迎，还因为它有许多吸引人的渐近性【Asymptotic Properties】（大样本属性）。在比较宽泛的条件下，最大似然估计是一致、高效和抽样分布为正态分布的，有着清晰的定义和易于计算的均值和方差。

如果对于任何的正数 $\varepsilon$，都有 $\lim_{n\to\infty} P(|\hat{\theta}_n - \theta| > \varepsilon) = 0$，那么基于容量为 $n$ 的样本的估计量 $\hat{\theta}_n$ 就是总体参数 $\theta$ 的一致估计量。最大似然估计量是一致的。这一结果说明，MLE 的估计值将随着样本量变得越来越大，而且越来越接近总体参数。而当样本量接近正无穷大时，估计值将收敛于总体参数。这一致性的结果是凭直觉获知的。让我们来想一想抛掷硬币问题。这枚硬币有未知的结果为面朝上的概率（$p$），共被抛掷了 $n$ 次。如果抛掷是独立的，在 $n$ 次抛掷中面朝上的次数 $Y$ 有二项分布。样本比例 $Y/n$ 是 $p$ 的一个估计量。我们不难理解，随着抛掷次数 $n$ 的增大，$p$ 的估计值将越来越接近它的实际值，而当抛掷次数接近正无限大的时候，估计值将收敛于 $p$ 的实际值。

估计量的效率由与估计量的估计值关联的变异性表明。一个总体参数 $\theta$ 也许有两个估计量 $\hat{\theta}_1$ 和 $\hat{\theta}_2$。这两个估计量代表两种不同的数学方法。给定一个样本 $n$，这两种方法将产生两个估计值。假设两组估计值都有一个等于 $\theta$ 的均值，这意味着 $\hat{\theta}_1$ 和 $\hat{\theta}_2$ 都是无偏的。那么，如果与 $\hat{\theta}_1$ 关联的方差小于与 $\hat{\theta}_2$ 关联的方差，我们就说 $\hat{\theta}_1$ 比 $\hat{\theta}_2$ 更有效。由 Rao-Cramér 得出的著名的结果给出了 $\theta$ 的每一个无偏估计量的下限。如果一个无偏估计量有等于 Rao-Cramér 的下限，那么我们便可知道，没有一个无偏估计量会有更小的方差，因而正在考虑之中的那个估计量便是无偏最小方差估计量。在相当一般的条件下，$\theta$ 的 MLE $\hat{\theta}$ 有一个均值为零，方差等于 Rao-Cramér 的下限的近似的正态分布。因此，至少近似的 $\hat{\theta}$ 是一个无偏的最小方差估计量。因此，在可以很好地利用收集的样本这一意义上，MLE 可以被认为是一个有效的估计量。

我们在这里要考虑的 MLE 的最后一个性质也许是最有用的。在最常规的情况下，我们可以证明，基于大样本的 MLE 有均值为 $\theta$ 和方差为 $[I(\theta)]^{-1}$ 的正态分布，在 $\theta$ 有单个参数时，有

$$I(\theta)^{-1} = \left(-E\left[\frac{\partial^2 \ln L(\theta)}{\partial \theta^2}\right]\right)^{-1}$$

在期望值可求得时，方差可以在 $\hat{\theta}$ 处评

估。实际上,对数似然的二次导数的期望值几乎是不可能得到的。在这种情况中,我们常常要计算的是经验信息矩阵,而非信息:

$$\left(-\frac{\partial^2 \ln L(\hat{\theta})}{\partial \hat{\theta}^2}\right)^{-1}$$

无论数据是否有正态分布,这个结果都是有效的。这一结果是回归分析中的 $z$-检验的基础,因此它在社会科学研究中备受青睐。

## 贝叶斯估计

近年来,贝叶斯估计在主流的统计学和生物统计学文献中用得越来越多。经典的估计量,如普通最小二乘估计量或最大似然估计量完全取决于样本。相反,贝叶斯估计既从样本也从先前的分布来获取信息。在有其他数据可资利用时,我们便可更新我们的先验分布【Prior Distribution】。贝叶斯估计整合我们先前有关研究问题的知识,因而可能更为主观。对大样本和常规模型而言,先前的分布通常会因为数据量而无从处理,而贝叶斯估计产生的估计值与最大似然估计并无二致。因此,贝叶斯估计并非最大似然估计的一般替代。

然而,至少在两种场合,贝叶斯估计可以起到很重要的作用。第一种场合是在观察数很少,但有关每个观察的信息却很充足时。有时宏观社会学研究便是这样。在一个样本是由国家组成的时候,样本中的国家数目必定有限,但有关每个国家的背景信息可能很多。在这种情况下,利用有关每一国家先前知识,也即利用数据之外的信息,必定会对分析有所帮助。第二种场合是在模型用经典的统计方法难以估计时。例子包括非线性的多水平模型。这些模型中的对数似然常为我们所涉足,使用无封闭表达式的多重积分。用在贝叶斯理论框架内开发的马尔科夫链蒙特卡洛法【Markov Chain Monte Carlo(MCMC) Methods】,先通过从要求的分布中抽取样本,绕开问题,然后再描述样本的平均数。还有一些在这里未曾讨论的其他估计法,例如,矩量估计法(method of moment estimation)和用于人口学数据的间接估计法。

——Guang Guo

(沈崇麟译校)

## 参考文献


Greene, W. H. (1997). *Econometric analysis* (3rd ed.). Upper Saddle River, NJ: Prentice-Hall.

Hogg, R. V., & Craig, A. T. (1978). *Introduction to mathematical statistics* (4th ed.). New York: Macmillan.

Lee, P. M. (1997). *Bayesian statistics: An introduction* (2nd ed.). London: Arnold.


# 估计量(Estimator)

估计量也称为(样本)统计量【Statistic】,它是一种规则、方法或程序,告诉我们能如何估计总体【Population】的数量,例如,总体均值、总体方差或总体相关【Correlation】系数。一般它都被表达为样本值的函数。在一给定的样本中,估计量所取的数值称为估计值。因为它的值会随样本的变换而变化,因此估计量是一个随机变量【Random Variable】,因而将会有一个概率【Probability】或抽样分布【Sampling Distribution】。在确认涉及总体值的估计值的信度时,估计量的抽样分布至关重要。所

有这一切可作如下解释。

令 $X$ 为一有概率密度函数【Probability Density Function, PDF】$f(X)$ 的随机变量。假定这一 PDF 有两个参数【Parameter】,$\mu_x$(总体均值)和 $\sigma_x^2$(总体方差)。基于来自这一 PDF 的 $n$ 个观察的一个随机抽样【Random Sampling】,假设我们想要估计均值 $\mu_x$。总体均值使用最为普遍的估计量是样本均值【Sample Mean】,其定义为 $\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$。因此,$\overline{X}$是一种规则或公式,它告诉我们求样本均值只要先将样本中的所有 $X$ 的值加起来,然后再除以样本观察数 $n$ 就可以了。因此,$\overline{X}$是 $\mu_x$ 的一个估计量(除了样本均值之外,样本众数【Mode】和中位数【Median】也是普遍使用的集中趋势【Central Tendency】的估计量)。

如果我们对估计总体的方差感兴趣,那么使用最为普遍的估计量就是样本方差【Sample Variance】,它以 $S_x^2$ 表示,其定义如下:

$$S_x^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2$$

根据这一公式,为了估计样本方差,先要从每一单个的 $X$ 值减去 $X$ 的均值,先取这些差的平方,再取这些平方差之和,然后再除以自由度【Degrees of Freedom】$(n-1)$。因此,样本方差被定义为真 $\sigma_x^2$ 的一个无偏估计量。

上面给出的计算样本均值和方差的公式给我们提供了一种叫作点估计值【Point Estimates】的估计值。这就是说,对于一个给定的样本,它们将给我们提供一个单一的(点)估计值。因为一个估计量是一个随机变量,所以它的值将随样本的变化而变化,就对它的真总体值而言,该点估计值的可靠程度究竟如何呢?在统计学中,一个点估计量的信度是用它的标准误【Standard Error】来测量的。因此,我们并不只是单单依赖点估计,相反,我们可以围绕点估计量构建一个区间,比如说,在一个点估量两侧的两个或三个标准误差之内构建一个区间,这样这个区间就会有一定的概率,譬如说,有 95%的概率包括真的参数值。因此建立起来的区间给我们提供了一种所谓的区间估计量。

已知 $X$ 的方差是 $\sigma_x^2$,统计理论可以证明$\overline{X}$的方差为:

$$\overline{X}=\frac{\sigma_x^2}{n}$$

因此,$\overline{X}$的标准误是它的方差的平方根,等于

$$\frac{\sigma_x}{\sqrt{n}}$$

统计理论也可以证明,如果观察数相当大,样本均值$\overline{X}$接近正态分布【Normal Distribution】,有均值 $\mu_x$ 和方差$\frac{\sigma_x^2}{n}$(这是以中心极限定理【Central Limit Theorem】为依据的)。如果随机变量 $X$ 最初来自一个正态的总体,那么不论样本的大小,$\overline{X}$都是有着规定的参数的正态分布。换言之,对于大样本

$$\overline{X}\sim N\left(\mu_x,\frac{\sigma_x^2}{n}\right)$$

$\overline{X}$是正态分布的,有规定的均值和方差。因此,样本均值的概率或抽样分布是正态的。

使用正态分布的一些性质,便可得到如下的总体均值区间

$$\overline{X} + 1.96\frac{\sigma_x}{\sqrt{n}}$$

它给我们提供了一个总体均值的95%的置信区间。如果我们建立100个这样的区间,那么将有95个有机会包含真的$\mu_x$。因此,前述区间是真均值的一个区间估计量。

求点估计量的方法有好几种,其中最为著名的是普通最小二乘【Ordinary Least Squares,OLS】法和最大似然估计【Maximum Likelihood Estimation,ML】法。就这两种方法而言,OLS法在数学上使用更简单。为了说明这一点,我们来看一下下面的多元回归【Multiple Regression】:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \cdots + \beta_k X_{ki} + u_i$$

式中的$Y$是因变量,带下标的各个$X$项都是解释变量,而$u_i$是随机误差项。为了估计这个模型,给定了数据,OLS具体步骤如下:重写前面的方程

$$u_i = Y_i - \beta_1 - \beta_2 X_{2i} - \beta_3 X_{3i} - \cdots - \beta_k X_{ki}$$

因此,$u_i$代表实际的$Y$值和它们来自多元回归的估计值之间的差(或误差)。OLS并不是最小化这些误差,而是最小化这些误差的平方。也就是最小化

$$\sum u_i^2 = \sum (Y_i - \beta_1 - \beta_2 X_{2i} - \beta_3 X_{3i} - \cdots - \beta_k X_{ki})^2$$

因为$Y$和$X$变量的值是由我们手头的样本提供的,因此前面的误差平方和将取决于$\beta$系数的取值。OLS试图找到这样一些$\beta$系数的值,它们将能使$\sum u_i^2$尽可能地小。实际计算$k$个$\beta$系数的代数公式是相当冗长的。使用矩阵代数,OLS估计量可以被写作:

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

式中的$\hat{\beta}$是一个$k$个$\beta$系数的($k\times1$)向量,$\mathbf{X}$是一个($n\times k$)的数据(或观察)矩阵,而$\mathbf{y}$则是一个因变量上的观察的($n\times1$)向量;符号“^”叫作帽子或者盖子,表示一个估计量。尽管$\beta$(向量)是未知的,但却是固定的,$\hat{\beta}$(向量)是随机的,因为它取决于$Y$和$X$变量的样本特定值。我们可以进一步证明误差的方差的OLS估计量是由下面的式子给出的:

$$\hat{\sigma}_u^2 = \frac{\sum \hat{u}_i^2}{n-k}$$

式中的$\hat{u}_i$是估计的误差项,而式中的($n-k$)代表自由度数,等于样本量$n$减去模型中待估参数的数目。

在对竞争的估计法进行选择时,我们可以考量从各种方法求得的估计量的统计特征。普遍使用的统计特征包括线性、无偏性【Unbiasedness】、最小方差、效率【Efficiency】、充足、渐近(参见渐近性【Asymptotic Properties】)效率和一致性。并非每一个估计方法都会满足所有这些统计特征,特别是在样本量比较小的时候,情况更加如此。

例如,如果假定上面提到的多元回归满足经典线性回归【Linear Regression】模型的假定,它可以给我们提供的OLS估计量是线性的、无偏的,且在一组线性无偏估计量中有最小的方差。常被称为最佳线性无偏估计量【Best Linear Unbiased Estimators,BLUE】。(经典线性回归模型假定,在重复抽样中解释变量的取值是固定的,误差项$u_i$是不相关的,$u_i$的误差的方差$\sigma_u^2$是常数或同方差性【Homoskedastic】的)。我们可进一步证明,上面给出的误差方差的OLS估计量是一个真$\sigma_u^2$的无偏估计量。

如果除了满足这些经典的线性模型的假定之外,它还假定误差项 $u_i$ 是正态分布的话,那么我们就可以证明 $\beta$ 系数的 ML 估计量等同于 OLS 估计量。然而,误差 $u_i$ 的方差的 ML 估计量是:

$$\sigma_u^2 = \frac{\sum \hat{u}_i^2}{n}$$

它不同于上面给出的 OLS 估计量,误差的方差的 OLS 估计量是无偏的,这意味着 ML 的误差的方差估计量是有偏的。两个估计量之间的差异是因为 OLS 估计量将自由度也考虑了进去,而 ML 估计量却没有这一事实。当然,随着样本量 $n$ 的增加,这两个估计量都将收敛。换言之,误差方差的 ML 估计量将渐近无偏。

——Damodar N. Gujarati

(沈崇麟译校)

## 参考文献

Gujarati, D. N. (2002). *Basic econometrics* (4th ed.). New York: McGraw-Hill.

Kendall, M. G., & Buckland, W. R. (1971). *A dictionary of statistical terms* (3rd ed.). Edinburgh: Oliver & Boyd.

Vogt, P. W. (1993). *Dictionary of statistics and methodology: A nontechnical guide for the social sciences.* Newbury Park, CA: Sage.

# Eta($\eta$)

在社会科学中,eta(希腊字母 $\eta$)一词可能是最著名的,因为它为海斯(Hays, 1988)和基尔克(Kirk, 1955)所用。这一词在基础的统计训练或理论语境中,并没有任何特别的含义。例如,在肯德尔和巴克兰(Kendall & Buckland, 1982)编撰的统计词典中就找不到它。搜索在《当代统计学索引》(*Current Index to Statistics*)上的所有从1980年到2002年出版的杂志文章,基于关键词 $\eta$(或 $\eta$ 方)的论文数为零。海斯和基尔克在方差分析【Analysis of Variance, ANOVA】语境中使用这个词,但是几篇近年来发表的关注这一专题的文章却并未提及 $\eta$。有些人用这一词来指效应量量度,它是学生 $t$ 统计量的函数,并在结构方程模型中有一定的作用。有时 $\eta$ 方在回归中被用来指复决定系数【Coefficient of multiple Determination】,但它更为普遍的是被标为了 $R$ 方【$R$-Squared】。

我们来看一个标准的方差分析模型,该模型要对所有 $J$ 个独立组的假定有正态性和有着公共方差的均值进行比较。$\eta$ 方这一词指一种旨在反映自变量和因变量之间任何非线性相关的关联量度。例如,有某种正在进行临床试验的治疗抑郁症的药物,试验使用5种不同的剂量,为阐述的方便,假定这5种剂量分别为10,20,30,40和50个单位的药物。我们感兴趣的是反映抑郁程度的某种结果量度的均值。如果5个组的均值不同,有一种可能性是随着药物剂量的增大,均值将以线性方式增长。另一种可能性则是一种非线性关联,例如,在这样一种关联中,在药物的剂量从10增加到20和从20增加到30时,均值则增加同样的量,但是之后,在剂量从30增加到40和从40增加到50时,均值却下降了。$\eta$ 方一词则指一种旨在反映这种或任何种类非线性关联的相关形式。总之,它的目的是反映任何一种对线性趋势的背离。

在 ANOVA 中，两个标准的数量是反映样本均值之间的变差的组间平方和（SSBG）和总平方和（SSTOT）。$\eta$ 的估计值等于 SSBG 的平方根除以 SSTOT。还有一种反映基于均值的、特定的、线性对比的、线性关联的数量；它在本质上是皮尔逊相关系数 $r$ 的一个变异。如果的确存在这种线性关联，那么 $\eta$ 和 $r$ 应该具有相似的值，这样一种检查非线性关联的方法便可以对这两个数量进行比较。更为详细的介绍请参见基尔克的有关著作（Kirk，1995，pp.197-198）。

——Rand R. Wilcox

（沈崇麟译校）

## 参考文献

Hays，W.L.（1988）.*Statistics.* Fort Worth，TX：Holt，Rinehart and Winston.

Kendall，M. G.，& Buckland，W. R（1982）. *A dictionary of statistical terms.* New York：Longman.

Kirk，R. E.（1995）. *Experimental design.* Pacific Grove，CA：Brooks/Cole.

# 伦理准则（Ethical Codes）

伦理准则详述了一个特定群体的成员的合适行为。本文集中讨论的是社会科学的伦理准则。这些准则是有争议的。它们忽视了一个专业人士最起码的应该是道德的职责，因而很少能真正作为做实际工作的专业人士的指南。准则证明了专业人士意识到了自己对社会有超越经济利己主义，或常规道德标准的责任。它帮助有道德的专业人士们对抗妥协的压力，使得其他专业人士对恶劣工作的反对得以合法化，同时改善专业人士的声誉和工作环境。职业团队可能利用伦理准则来谴责或驱逐某些行为不当的成员。对伦理准则的违反尽管本身不会受到法律的制裁，但在某些法律诉讼中可以作为证据采纳。伦理准则的作用要放在历史语境下来理解。

## 历史

直到 1800 年左右，大家都认为专业人士是不需要书面规则来约束行为的绅士。很少有迹象表明一个专业人士的不名誉行为会导致法律诉讼、文字战，甚至直接的决斗。1792 年，当曼彻斯特医院的医生进行一场文字斗争，并因此在流行病爆发期间罢工时，这类特殊标准的不当性得到了大家的确认。愤怒的医院管理层要求珀西瓦尔（Percival）医生起草规则，防止未来对职业道德的损害。珀西瓦尔医生的准则公布于 1794 年，列示了医生的明确职责，申明了普遍道德标准的重要性，并建立了医生对官方权威的道德独立性，以对抗他们降低医疗服务标准的压力。1847 年，美国医学协会成为第一家接受珀西瓦尔医生准则的全国性专业协会，并将那些准则称为"伦理准则"。

## 社会科学中伦理准则的出现和发展

1953 年，美国心理学协会公布了第一部社会科学伦理准则，并采用了现在熟悉的行为实践：召集一个委员会，收集对协会成员遇到的道德困境的描述，并通过协商同意的方式发展出了一些恰当的规则。现在，社会科学的每一个分支都有伦理准则，每一个分支的子领域也同样建立了伦理准则，比如

舆论研究、临床催眠等。社会的(例如,对隐私的强调)、规章制度方面的(例如,机构审查委员会)、职业角色的和研究主题的变化都需要对准则进行经常的回顾检查和修改。典型的做法是,一个专业协会的伦理委员会会出版描述出现问题的文章,提出解决这些问题的建议,并寻求会员的同意。提出的准则草案将会被反复修改,直到会员形成一致意见后才公布出版。

## 使用准则

由于伦理准则是用来回应预期出现的伦理冲突的,因此对伦理准则的理解和解释最好放在真实生活的伦理困境语境中。伦理两难的个案研究对于教授伦理准则的使用是非常有效的工具,因此现在已成为每个学科的专业文献中的重要组成部分。

现在关于伦理准则已经积累了大量的文献。准则版本的收集、带注释的参考文献、个案汇编、研究指南、有效研究和伦理的方法论的整合准则,以及准则编写和准则使用教学指南,都可以在伊利诺伊理工学院(Illinois Institute of Technology)的职业伦理研究中心的网站中找到。伦理准则的档案,包括不同专业领域中准则的历史、建构、使用和案例,也都存放在这个职业伦理研究中心里。

——Joan E. Sieber

(郑晓娟译　沈崇麟校)

# 伦理原则(Ethical Principles)

伦理的话语可展现在不同的抽象层次。抽象层次由高到低包括伦理理论、原则、规则或规范、准则和判断。伦理原则的本质和作用最好能够与这一等级结构中的其他要素联系起来理解。

判断是针对特定行动的结论。以下是一个案例:

在一项我们观察抵制诱惑的研究中,我们不应该对5岁的小孩说,他们是否偷了我们丢下的钱,因为那个年纪的孩子都会马上去拿有吸引力的东西,而不会意识到他们是在做坏事。相反,我们应该和他们聊天,让他们感到自己的行为是没有问题的,并确保他们的父母了解儿童发展和我们研究的有关方面。这一非常具体和特别的判断,可以根据更抽象和普遍的规则、原则和理论被认为是正当的。

一条伦理准则明确了一个特定群体成员的合适行为(见伦理准则【Ethical Codes】)。这意味着,一条伦理准则是一个特定群体的一组规则或规范。该准则明确了群体对社会的特定责任,这一责任超越了普通的道德标准。例如,美国心理学协会的伦理准则不仅要求要作出恰当的事后报告,而且还要证明上面讨论的,对欺骗研究中年轻参与者的事后报告的判断和规范是正当的。该准则陈述道:

(a)心理学家应及时给参与者提供得到有关研究的性质、结果和结论的机会,并尝试纠正参与者可能会有的任何错误概念。

(b)如果科学或人道价值观证明有正当的理由,推迟或阻止这些信息的发布,心理学家应当采取合理的措施来降低损害的风险。(American Psychological Association, 1992, p.1204)

注意这一陈述比前面的伦理规则或规范更加宽泛。它引发了对有关准则的恰当

解释的判断。

伦理规则或规范通常是指那些非正式或没有形成书面文字，但一般都被理解为有关是非对错的一般性陈述。例如，研究者可能认为上面的伦理判断是正确的，他依据的规则或规范是："研究者应该向被研究者报告，但是不应该以一种会对被研究者及其家长造成伤害或担忧的方式来进行。"

一条伦理原则是一条更为广义的规则，例如"尊重研究参与者""公平""防害"或"增益"。它陈述理想的状态，忽视不同原则之间可能有的冲突，因而对于人们应如何实践这些理念并没有指导作用。例如，美国心理学协会像下面这样来陈述F原则：

> 心理学家应意识到对于自己工作和生活的社区及社会的职业和科学的责任。他们应用和公开自己的心理学知识的目的是为人类福祉作出贡献。心理学家关注并致力于减轻人类痛苦的根源。在从事研究过程中，他们力求增进人类福祉及推进心理科学的发展。心理学家尽力避免对他们工作的误用。(American Psychological Association, 1992, p.1600)

伦理理论是对原则和规范的系统性陈述。下面是对一些主要类型的伦理理论高度简化的描述：功利主义(结果)理论认为，个体应该去做任何会增加世界的快乐的事情。义务论伦理学(康德的版本)认为，个体有需要承担的特定责任，无论可能的结果如何，因为这些责任总是具有道德的正确性，除非被更高的原则(例如拯救生命)所代替。这些责任包括说真话，或者将个人看作结果而不是达成某种结果的手段。与此类似，一些责任主义者可能会认为，某些行为(如堕胎)总是错误的。有神论的伦理则认为，人应当做那些使上帝愉悦的事。

诚如这些简单的定义和例子所示，不同的伦理理论能够为不同的原则、规范或判断提供依据；伦理学家可能无法就什么样的行为才是道德上或伦理上正确的这一问题达成一致。此外，同样的理论和原则也可能会由不同的人或在不同情境下给出不同的解释。

不一致、例外和冲突——伦理理论、原则和规范总是有一些例外的情况。例如，对于伦理理论的例外情况，那些认为杀人总是错误的道义论者，可能也会认为存在一种像"正义战争"那样的正当杀人，那些要求进行真情告知的道义论者，可能允许为了拯救一个无辜的人而撒谎。我们常常会发现，伦理原则之间存在冲突或例外。伦理理论最重要的作用之一，就是帮助解决伦理原则之间的冲突。研究5岁儿童对诱惑的抵制能力的研究者，试图调和诚实报告和不对研究参与者造成伤害这两个方面的要求，他们向注重利益的功利主义原则求助，研究者有义务将研究的获益最大化并使可能的伤害最小化——为了正当的理由。

本词条的目的在于对一些伦理原则及其在话语层级中的位置做一次简要的介绍，其实不仅讨论伦理学的方式有许多种，而且我们可取的立场也不尽相同。那些看似难以理解的意见分歧常可通过对其背后的理论加以识别来理解。例如，那些不齿于研究实践中的隐瞒或欺骗的学者似乎都是一些道义论者。因此，那些未曾明示自己在科学讨论中所持关键立场观察他人行为的文化人类学家，其行为可能会被认为是错误的，因为他们把被研究者当作了解人类行为这一目的的手段，而不是把他们本人作为终极目的。另一方面，功利主义者则可能认为，人类最好还是要能理解自己，即使这要以对个别被研究者做一些微小欺瞒为代价，目的正当即手段正当。由《贝尔蒙报告》(*Belmont Report*)提出，并经"联邦保护研究中的人类被研究对象条例(Federal Regulations for Protection of Human Subjects of Research)"颁布的伦理原则就是功利主

义和道义论伦理的混合产物。

研究者需要解决的最为困难的伦理问题之一就是,当他们需要进行一些有风险性的研究以便获得一些非常重要的知识时,那个研究需要诉诸功利主义原则,因而对道义论原则要有所妥协。例如,对于未成年人注射药物的研究可能对社会有非常大的好处,因为这样的研究可能有助于增强社会未来控制药物使用方面的能力。但是,研究者不可能获得这些未成年人的父母关于开展这项研究的同意;父母可能根本不知道自己的孩子有药物使用问题,或者孩子依赖药物使用维持生存。伦理原则有一个好处,是其比较宽泛和普遍,因此能够为我们考虑各方面因素提供灵活的指导意见。同样,正是由于其宽泛和普遍的特点,伦理原则在一些具体问题上也表现出了模糊不清,很容易被误解。带着这样的思维限制,我们再次回到由美国保护生物医学和行为研究中的人类被研究者全国委员会(the U. S. National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research)(全国性的委员会)1978 年所建立的贝尔蒙原则(Belmont Principles)。该委员会的部分职责是为指导人类研究设立基本的伦理原则。经过慎重考虑,它建立了三个伦理原则,支持科学研究中六个广泛达成一致的规范。三个伦理原则包括:

①有益:要最大限度地为科学、人类和参与研究的个体争取良好结果,同时避免或者将不必要的风险、损害及错误最小化。

②尊重:保护(有自主能力的)人的自主性,将每个个体作为人给予礼貌对待和尊重,包括那些没有自主性的个体(例如婴儿、弱智人士或受衰老折磨的人)。

③公正:确保程序的合理、非压榨以及考虑周全,并确保程序的公平执行,包括在不同个体和群体间公平分配成本和收益(例如,那些承担了研究风险的研究参与者也应当从中收益)。

科学行为的六个规范如下,括号中的字母表示每个规范的建立是基于哪个伦理原则的:

①有效的研究设计。有效研究充分考虑了相关的既往研究、方法和理论。无效的研究会产生无效的结果、对被研究者的不尊重和资源的浪费。(A,B)

②研究者技能。研究者必须有足够的能力来有效地开展研究。(A,B)

③确知后果。研究者应当明确了解研究的风险和收益。同时应调整研究程序以便尊重隐私,确保保密,使收益最大化、风险最小化以及达成其他目标。(A,B,C)

④被研究者的选择。被研究者的选择必须要适合研究本身,要对受益的总体具有代表性,且数量适当。(A,B,C)

⑤自愿知情同意。研究应获得被研究者的自愿知情同意。(A,B,C)

⑥对伤害的补偿。研究者应对被研究者的伤害负责。研究者应告知被研究者如果造成伤害是否会给予补偿。(A,B,C)

贝尔蒙原则体现了美国文化中的个人主义和实用主义特征。如果是更具家长制或公有制的文化,可能不会像贝尔蒙原则那样更多强调对自主性的尊重,而会更多地强调受益性原则。一个更具有神权色彩的文化,可能不鼓励对人类进行研究,因为他们认为所有的行为都必须遵从上帝的意志。除了这些比较极端的道德观念之外,全国委员会可能还选择了一些在某种程度上持不同观点的基本原则。许多学者已经发现,现在对自主性的强调可能会使研究参与者面对一些无力作出或不愿作出的决定。因而,许多研究参与者在根本不理解(甚至没有阅读)知情同意书的情况下,作出了同意或拒绝参与研究的决定。如果我们的研究者在不损害参与者利益的情况下,能够表现出更

具关爱或更专业的责任,这样是不是更符合伦理原则呢?这个问题的答案似乎并不容易给出。

——Joan E. Sieber

(郑晓娟译　沈崇麟校)

## 参考文献


American Psychological Association. (1992). Ethical principles of psychologists and code of conduct. *American Psychologist*, *47*, 1597-1611.

National Commission for Protection of Human Subjects of Biomedical and Behavioral Research. (1978). *The Belmont report: Ethical principles and guidelines for the protection of human subjects of research* (DHEW Publication No. (OS) 78-0012). Washington, DC: Government Printing Office.

Solomon, W.D. (1978). Ethics: Rules and principles. In W. T. Reich (Ed.), *Encyclopedia of bioethics*. New York: Free Press.


## Ethnograph

定性数据分析软件是一种设计用于计算机辅助的定性数据分析软件。正因如此,它能对定性数据进行编码和检索。此外,它也支持扎根理论【Grounded Theory】的实践,如作备忘录,撰写备忘录(the generation of memos)。

——Alan Bryman

(沈崇麟译校)

## 民族志内容分析(Ethnographic Content Analysis)

民族志内容分析(ECA)指的是一种将定位寻找、识别、检索和分析文献的相关性、重要性和意义融为一体的方法、程序和技术(Altheide, 1987, 1996)。其重点是发现和描述,包括搜寻背景、潜在的意义、模式和过程,而不只是两个或多个变量之间的数量或数字关系(Altheide, 1996)。

文献的定义是任何可被记录和/或被检索的,用来进行分析的符号表示和语意。随着记录技术的提高,文献分析的技术也会随之发展,变得更易于存取。这些技术包括印刷和电子媒体、录音带、视觉资料(例如照片、家庭录影带)、衣着/时尚、互联网材料、信息库(例如 Lexis/Nexis)和田野日志等。

ECA 或定性文献分析涉及对源于信息库(包括研究者所开发的信息库,例如田野笔记【Fieldnotes】)的文献的自然和理论抽样(Glaser & Strauss, 1967),以及更为系统的分析方案的开发,然后通过比较来廓清主题、框架和话语。举例来说,如果有人对研究电视暴力感兴趣,那么他们感兴趣的并非具有社会意义的暴力行为本身,而是暴力行为如何与行动步骤,或者作为娱乐重点(entertainment emphasis)部分的情节(例如,"坏人被好人击毙,以获得正义")关联。或者,情节也可以是暴力的使用多少都与勇敢、机智、技巧或(当然)性有关联。后者是具体情节里反复出现的主题和一般信息。而我们的目标就是要探询行为和事件是如何被放置在情境之中的,以及表现的主题、框架和话语是什么。基本步骤包括:

- 追问一个具体的调查问题。
- 对信息源(例如对报纸或电视台的民

族志研究）的过程和情境逐渐熟悉起来，探索可能的信息（或者文献）来源。

- 对相关文献的几个（6～10个）例子逐渐熟悉起来，选择一个分析单位（如每篇文章），认识到分析单位可能会发生变化。
- 列出几个项目或类别（变量）以指导数据收集和实验方案拟订（数据收集单）。
- 用从几个文献中收集来的数据来检验实验方案。
- 修改方案，再选择几个案例来进一步改进实验方案。

动态地使用 ECA 牵涉“追踪话语”，或者在一段时间内跨不同的主题和不同的新闻媒体来追踪某些话题、词汇、主题和框架。最初的显性编码（manifest coding）结合了自然编码和理论抽样，以监测范围和重点的跨时间和跨主题的变化。比如说，在对于“恐惧”的研究中（Altheide，2002），制订出一个方案来获取有关恐惧的日期、地点、作者、格式、话题、资料来源、主题、重点和语法运用（“恐惧”作为名词、动词和副词）。通过理论抽样和用以描绘模式和主题重点的持续比较法【Constant Comparison】，明确使用“恐惧”这个词的情境。用一个词语处理和定性数据分析程序 NUD＊IST，可以定性和定量地进行材料的列举、绘图和分析。

——David L. Altheide

（杨可译　沈崇麟校）

## 参考文献

Altheide，D.L.（1987）.Ethnographic content analysis. *Qualitative Sociology*，*10*，65-77.

Altheide，D. L.（1996）. *Qualitative media analysis.* Newbury Park，CA：Sage.

Altheide，D.L.（2002）. *Creating fear*：*News and the construction of crisis.* Hawthorne，NY：Aldine de Gruyter.

Glaser，B.G.，& Strauss，A.L.（1967）. *Discovery of grounded theory*：*Strategies for qualitative research.* Chicago：Aldine.

# 民族志的实在论传统（Ethnographic Realism）

民族志的实在论认为民族志【Ethnography】可以而且应该努力把社会文化的结构、过程和情境当作独立于研究者的存在记录下来。

纵观民族志的发展史，它大多时候都致力于唯实论【Realism】和以实在论为前提的调查方法。比如说，人类学家长期主张，为了正确地理解其他社会的文化和社会组织，必须通过近距离的参与观察【Participant Observation】来研究。在20世纪的美国社会学中，布卢默·赫伯特和其他人推荐的也差不多是同样的方法，强调要捕捉人类社会生活的阐释和过程特点，他们对未能做到这一点的方法提出了批评。1960年代末，大卫·马茨把这种实在论——有时也称为自然主义【Naturalism】——概括为保存被研究现象的真实性质的责任。

不过，民族志的实在论在1980和1990年代遭到了挑战。批评家们对提供有关社会现象“本质”的描述是否可能提出了质疑，有时批评家们否认这些现象独立于调查的过程而存在。这种批评在一定程度上发端于民族志本身。一般来说，民族志学者所描写的人们总是在积极地理解世界，由此也发展出了独特的文化。如果这种方法被反思性地用于民族志叙述自身，那么可能会得出这样的结论，即这些叙述是反映研究者的社会文化身份的创造性建构，而非客观的呈

现。在某些大陆哲学流派的影响下，这种结论更加受到鼓励，这些大陆学派主要是现象学【Phenomenology】、诠释学【Hermeneutics】和后结构主义【Poststructuralism】（尽管这些立场并不都含有认识论的相对主义【Relativism】或怀疑主义）。

另一种重要的批评指出，民族志叙述常常与实在论小说一样依赖于相同的文学手段。这些批评家坚持认为，没有一种写作形式只是在表现：所有的语言运用都是构成的和述行的，因此作家应该都是这个样子。

伴随着对民族志实在论的认识论和修辞基础的批评，对它的伦理和政治上的批评也一直没有停止过。这种批评受女权主义以及1960年代激进左派挑战社会学的启发，提出实在论的民族志相当于一种窥阴癖（voyeurism），和/或一种监督的手段，被用来控制那些处于从属社会地位的人。

在推出这些论点的同时，有些民族志实在论的批评者试图发展一些新的写作形式，其目的是颠覆民族志实在论给人们留下的"民族志文本可以提供有关世界的客观叙述"的印象，并展现他们自己建构的、不确定的特点，和/或让那些被主流叙述视为"沉默"的人们发声（Denzin，1997）。

也有人一直在为民族志实在论所受到的批评进行辩护。这种怀疑论和相对论主张的自损特点一直在不断地重复。不仅如此，在有时被归为属于过去的民族志学者的朴素实在论和更复杂的形式之间也已作了区分。而较为复杂的实在论考虑了认识论的难题和所有研究的社会情境特性，但没有否认社会科学知识的可能性，也并没有坚持认为民族志总是不可避免地要与支持或抵制现状的政治斗争纠缠在一起（Hammersley，1992）。

——Martyn Hammersley

（杨可译　沈崇麟校）

## 参考文献


Atkinson, P.(1990). *The ethnographic imagination: Textual constructions of reality*. London: Routledge.

Denzin, N. K. (1997). *Interpretive ethnography*. Thousand Oaks, CA: Sage.

Hammersley, M.(1992). *What's wrong with ethnography?* London: Routledge.

Matza, D. (1969). *Becoming deviant*. Englewood Cliffs, NJ: Prentice-Hall.


## 民族志故事（Ethnographic Tales）

民族志【Ethnography】是一种人类学使用多年的数据收集方法，它提供了有关个体和群体的丰富、详细的信息。作为一种定性研究方法，民族志一般都包括参与观察【Participant Observation】或非参与观察【Non-participant Observation】，和/或深度访谈【Indepth Interview】，让研究者得以对感兴趣的人群进行综合考察。

好的民族志叙述提供了"浓描"，或者说一种对社会情境的细节描述，分享了一个特定群体的意义（Geertz，2000）。正因如此，民族志研究才会产生大量的数据——不夸张地说是成百上千页的田野笔记【Fieldnotes】、分析性的备忘录【Memos】，和/或访谈的文字稿。在一定程度上，正是因为民族志方法收集的信息量如此巨大，以致社会科学家从这些数据中建构出来的故事可能也会有许多不同的形式。尽管民族志学者共有的目标都是为一个群体或社会现象提供深度观察，但他们却可能会选择若干种

方法来呈现他们的发现。

范·曼内斯(Van Maanen,1988)在他的著作《田野的故事:论文字民族志》(*Tales of the field*: *On writing ethnography*)中认定了三种主要的民族志故事:实在论的、忏悔式的和印象派的。实在论的民族志故事目的是提供一个有关被研究群体的客观视角。因此,实在论的故事类似于更加传统的民族志叙述,读者可以从中得到一个对被研究的个体或群体清晰、简洁的描述。他们不会感受到作者对这个环境的反应、作者所感觉到的其自身对研究发现的影响,也无法从中体察作者从"田野"回来之后的反思。

另一方面,忏悔式的故事则让读者得以了解作者的田野经验是怎样的。尽管有些人会因为这种故事的主观性而认为它不那么科学,因此是有缺陷的,但在评价研究的影响力时,忏悔式故事仍不失为一种有价值的手段,因为它们可以让读者了解作者的方法和偏见,以及研究进行时在方法论上得到的经验教训。反身性【Reflexivity】和对研究过程的披露对女权主义的定性研究者来说一直都很重要,他们的目标就是在提供有意义的发现之外,还要阐明科学的研究方法。

印象派的故事是对非典型现象的叙述。印象派故事的目标是用对独特发现的朴实描述来震撼和吸引读者。在这种故事中,研究者不提供对发现的解读,而是呈上非同寻常的故事,让故事自己来说话。

范·曼嫩的民族志故事类型不是互斥的类型,人们可以在一个叙事中展现两种或所有的故事类型。正如米勒(Miller,1998)在她对于无家可归者的施粥点的研究中所给出的解释,在作者反思和分享其发现时,这些故事可能会有点出人意料地浮现出来。此外,民族志和其他定性方法的一个首要目标就是要让边缘群体发声(Ragin,1994)。让其经验没有得到广泛认可的人们发声,而且要为研究发现提供合理的解释,可能也需要使用不同的民族志故事。

——Desirée Ciambrone

(杨可译　沈崇麟校)

## 参考文献

Geertz, C. (2000). *The interpretation of cultures*: *Selected essays*. New York: Basic Books.

Miller, D. (1998). Writing and retelling multiple ethnographic tales of a soup kitchen for the homeless. *Qualitative Inquiry*, *4*, 469-492.

Ragin, C. C. (1994). *Constructing social research*. Thousand Oaks, CA: Pine Forge Press.

Van Maanen, J. (1988). *Tales of the field*: *On writing ethnography*. Chicago: The University of Chicago Press.

## 民族志(Ethnography)

民族志是描写一个群体或文化的艺术和科学。它的长处在于描述。民族志从特定文化的本土人的视角给读者提供了一幅正在进行的事情的详细图像。对于一些导致特定结果的项目实施过程,它还能为我们提供对这些过程的洞见。它是涉及田野工作的主要方法学手段之一。民族志学者将自己沉浸在某个文化中,以便在其自然环境中观察和记录人们的行为。用于观察的时间可能从几个月到几年不等,必须要有足够的时间来观察一段时间内的行为模式。这是一种形式的信度【Reliability】。

民族志学者试图在田野中得出局内人或主位的(emic)现实观(参见主位/客位之

分【Emic/Etic Distinction】)。民族志学者可能同意,也可能不同意这位局内人的观点,或者认可它是对现实的客观观点。民族志的目标就是不做主观判断,它承认一个人对现实的认识,无论其是否具有科学价值,都会对现实有所影响。民族志学者承认,根据人们在社会中的不同角色,存在多重现实或者多种现实观。比如说,一位中上阶层的老年妇女和经济社会地位较低的美国黑人男性,对同一位警官的看法可能就不同。民族志学者记录了多种多样、时常互相矛盾的主位或局内人观点。大多数民族志学者以这样的知识为基础,并采用外部或客位的(etic)观点来解释这些现实的主位观之间的关系。

民族志比其他形式的研究更花时间,因为需要田野工作——要和人们一起在其生活的自然环境中共度一段时间。民族志学者在田野待上一段时间,以便把他们收集的数据置于情境之中。花在田野工作上的时间长短可能会影响研究者的适应,但花在田野上的时间和数据的质量直接相关。有讽刺意味的是,更高的效率常常却是没有效率:花费更多的时间在田野上,将会在调查完成之后节省为修正理解不充分的问题所要花费的时间。

这种沉浸在其他人生活中的民族志,使民族志学者得以更准确地解读人们的行为。这被称为文化阐释。克利福德·格尔茨(Geertz,1975)借用了吉尔伯特·里莱的浓描【Thick Description】的概念,给出了一部民族志的经典例子。这就是使眼色或眨眼睛这一例子,而正是这一例子告诉了我们,如何在民族志解读中捕捉文化阐释的意义。眼皮快速闭上又睁开,同样的机械运动在灯光昏暗、烟雾缭绕的酒吧可能是使眼色,而在满是灰尘的家居修理店就是眨眼睛。定义同样的机械行为应如何解读的正是这种文化情境(Fetterman,1998)。

## 田野工作

田野工作是民族志最重要的特点。民族志学者通常会采用一种自然主义的路数,在人们自己的环境中来研究他们。最常见的技术是一种分层判断样本(stratified judgmental sample),民族志学者依靠自身判断,根据其研究的问题,从该文化的不同类别或角色中选取最适合的成员来进行研究。问题通常都是非正式的、开放的,也没什么威胁。

田野工作有很多阶段。第一个阶段包括(根据研究计划)进入田野。一旦被引入了社区,民族志学者通常就要从一种调查了解基本情况的阶段入手,这包括了解语言、历史、政治、经济等基本情况。调查之后的阶段的主要特点就是不断的资料收集、理论概括、检验假设和对信息进行交叉检验。换句话说,民族志学者要花大量时间来推测各种关系和行为模式。观察和访谈被用于证实、反对或扩充有关情境的最佳推测。对于理解某种文化或群体而言,这种方法有助于打下一个扎实的基础。

田野工作后面的阶段关注点一般都比较狭窄和精确。对批评概念和主题的探讨往往更为深刻(但常常不是很宽广)。这时常会有时间压力,要找到缺失的信息、弥补空白,以及与研究的批判性发现关联最为紧密的矛盾之处,或者与研究关联的洞察。

## 参与观察

参与观察【Participant Observation】最能体现大多数民族志工作的特点,因而对有效的田野工作来说至关重要。参与观察把参与被研究人群的生活和保持职业距离结合起来,让研究者得以有充分的观察并记录数据。长期驻扎在当地可以帮助民族志学者内化该群体基本的信仰、恐惧、希望和期待。

## 概念

虽然民族志学者带着开放的态度进入田野,但并非脑中空空。民族志学者有指导其观察和阐释传统的概念。指导其工作的几个最重要的概念包括文化、整体的视角、情境化、主位/客位视角,以及非判断定位。

### 文化

文化是指导实践的最宽泛的民族志概念。文化包含了那些体现其人群特征的观念、信仰、知识和行为。文化的概念可以帮助民族志学者寻找某个群体内部合理的、有凝聚力的模式。每种文化都有一个可识别的价值体系,它形塑了人们的行为。发现某个群体背后的价值观就像解开人类基因组的奥秘一样。价值观构成了一种社会的DNA,有助于研究者理解和有可能预测一个群体内人的行为。

### 整体观

民族志学者在研究中采取了一种整体的观点,试图尽可能多地描述某种文化或社会群体,包括历史、宗教、政治和环境。没有研究可以抓住整个文化;但是,这个定位促使民族志学者进一步探究文化的深度和广度,并在更广大的社会和经济背景中来观察每个事件。这个概念透镜使民族志学者得以了解那些处于边缘地带的相互关系和包含的多重性,而这些东西常常为那种比较专门和狭窄的路数所忽视或忽略。

### 情境化

数据情境化指的是把观察置于更大的视角之下。一个事件或行为只有放在情境之中才能得到有意义的阐释。孤立的、脱离情境的行为或事件很容易被误读。更为重要的是,把事件或行为放置在其自身环境或一系列条件之下,能避免剥掉其多重意义。

### 主位/客位视角

主位的或局内人的视角是大多数民族志研究的核心。民族志学者在局内人观点的基础上建立自己对某个群体或文化的知识,而不是从一个事物如何运作的一组先验假定出发来开展工作的。在汇集了许多局内人观点之后,民族志学者通常也会采用客位的或外部的社会科学立场。其目标就是要弄清局内人是如何相互关联、交流和作为一个更大的系统或文化的一部分发挥作用的。

### 非判断定位

非判断定位有助于防止民族志学者对其观察的事物作出不合适、不必要的价值判断。非判断定位要求民族志学者停止对任何特定文化实践作个人的评价。保持非判断定位就好比人们在看电影或读书时停止怀疑一样——我们要接受可能显然不合逻辑或不可置信的一套环境,以允许作者展开一个引人入胜的故事。

我们都有个人的信仰、偏见和个人品位。民族志学者通过将偏见外显和力求公正地看待他文化来提防比较明显的偏见。民族中心主义的行为——把一种文化的价值观和规范强加于另一个民族,带有一种文化高于另一种文化这一假定——在民族志中是一个致命的错误。

许多其他的概念也形塑或指导着田野中的民族志学者,包括文化间和文化内的多样性、结构和功能、象征和仪式、微观和宏观层面的研究,以及操作主义。这些概念就像理论一样指导着民族志学者的实践。

## 方法

民族志学者是一种人性化的工具。人力工具依赖它所有的感觉、思想和感情,是最敏感和有领悟力的数据收集工具。不过,这个工具收集的数据可能是主观的、有误导性的。民族志方法和技术有助于指导民族志学者通过漫无边际的个人观察,进而确定令人迷惑的,各种各样的形成一个社会情境的事件和行动,并对之进行精确的分类。

尽管田野工作和参与观察代表了绝大多数人类学工作,不过民族志学者还是要依赖各种各样特别的工具,包括访谈、抽样调查和非介入性测量【Unobtrusive Measures】。

### 访谈

访谈是民族志学者最重要的数据收集技术。各种访谈可以解释民族志学者所见和所体验的内容,并将其放置在更大的情境之中。一般的访谈类型包括结构化访谈、半结构化访谈、非正式访谈和回溯式访谈。

结构化访谈和半结构化访谈是为了比较和呈现——比较各种回答,并将其放置在共同的群体信仰和主题的背景之中。非正式的访谈不同于谈话,但它一般常与某种谈话结合起来,形成了一种谈话和内嵌问题的混合体。问题多是从谈话中浮现出来的。有时候这些问题是偶然发现的,源自参与者的评论。民族志学者采用回溯式访谈来重构过去,请知情人来回忆个人的历史信息。

问题的类型。在一个访谈中就有许多种问题。抽样调查的问题或者观光旅行的问题就是设计出来以获取有关参与者或本地人世界观的广阔图景的。抽样调查的问题也有助于定义一项研究的边界。一旦抽样调查的问题揭示出某个有意义的类别,对这个类别的具体问题就变成最有用的了。具体问题进一步探测已确定的类别的意义或行动。抽样调查问题形成和报告总体性的理解,而具体问题则将这种理解加以精炼和扩展。结构和属性问题——具体问题的子类——常常用来进一步理清主位的观点。

此外,民族志学者在田野工作中也采用开放式问题和封闭式问题。开放式问题允许参与者来解读它。封闭式问题有有限的几个可供选择的选项,有助于厘清观点和量化行为模式。

### 关键行动者和生活史

有些人比别人更善于言谈,更具有文化敏感度。这些人可以成为很好的"关键行动者",作为民族志学者重要的信息来源。他们能提供详细的历史数据、有关当前人际关系的知识,对日常生活的微妙之处也了如指掌。尽管民族志学者试图跟尽可能多的人交谈,时间总是一个需要考虑的要素。因此,传统上人类学家非常依赖群体中的一两个人。关键行动者常常能给民族志学者提供丰富、详细的自传式描述,或者能确认群体中哪些成员值得进行这样深入的研究。这些生活史常常能为我们带来一幅描写被研究文化的完整的画面。

### 抽样调查

抽样调查是大规模数据收集的有效方式,也是处理代表性问题的好办法。抽样调查是为成千上万人号脉的唯一的办法(开展抽样调查的指南可参见 Fowler,1988,Fink & Kosecoff,1988)。除了抽样调查,民族志学者还经常使用非介入性测量。

### 非介入性测量

非介入性测量依靠实物证据来作出社会文化推论。某个露出地面的东西,或者说是经济或心理条件的物理表现,可能是涂鸦或者烧毁的建筑。不依靠任何人类互动就可以作出最初的推断。不过,这些线索本身可能是误导性的。还需要交叉检验和搜集其他的数据。

其他的非介入性测量包括对档案文献、电子邮件、学校记录、人口普查记录和财政预算的研究。人际距离学有助于区分亲密的家庭关系和疏远的、事务性的关系。它有助于我们阐释一种社会情况,即有时候人们的个人空间在许可或未许可的情况下就有人进入了。人体动作学或肢体语言也常常可以表示情绪和性情(Birdwhistell,1970;Hall,1974)。

## 设备

记事本、计算机、录音带和照相机——所有的民族志工具——都只是人力工具的延伸,支持和提升记忆、能力和视野。这些有用的设备可以捕捉丰富的信息和民族志经验的感觉,便利民族志工作,帮助组织和分析数据。

民族志学者最常见的工具就是纸和笔。用这两样工具,田野工作者可以在每次访谈中或访谈后记笔记,画出某个区域的物理布局,描画某个组织机构图或非正式社会网络图。录音机能在民族志学者保持自然的对话语流时有效地逐字逐句获取大段的引语,这是完善田野工作必不可少的。此外,访谈的文字稿【Transcription】尽管很花时间,但却是贴近数据的一种方法,研究者对讲话的模式、当地的参考文献和语调的抑扬顿挫更为熟悉。

照相机让民族志学者得以对特定行为进行图片记录。这是表面效度【Face Validity】的一种形式。图片也可以被用于投射技术;和关键行动者分享图片可能会引发他们对其行为作出深刻思考。在数据分析和写作的阶段,照片也可以补充和刺激民族志学者的记忆。数码相机现在是田野中的标准装置了。它可以让民族志学者立刻分享照片——在照相机的屏幕上、粘贴在电子邮件中,或是存在服务器上让团队或社区的成员(在许可后)来观看。

数码录像机也是田野里的标准配置。民族志学者用不太贵的软件来制作有关关键事件的电影。这些电影对民族志发现进行可视化的表现或再现(有关民族志电影制作指南,请参见 Heider,1976,Bellman & Jules-Rosette,1977)。

计算机在民族志中是必不可少的。笔记本电脑用于做田野访谈记录。诸如 NUD * IST、定性分析软件【Ethnograph】和 ATLAS.ti 等定性数据分析软件被用来给田野记录编码、分组和排序,并识别模式。在给赞助人或同事的报告中,PowerPoint 演示用于突出田野中的发现。

互联网包含了民族志工作中的几种有用的工具,包括电子邮件、搜索引擎以及诸如电话簿和地图等参考网页。此外,互联网上还有许多免费的或不太贵的在线调查项目。互联网上的视频会议是在田野中收集数据,或在离开田野之后与关键行动者保持正常联系的一种方法。这些工具能节约时间和金钱,因而它们都能给交流带来很大的便利(更多细节请参见 Fetterman,2002)。

## 写作

民族志要求每一阶段的写作都必须力求完美。研究计划、田野日志、备忘录、中期报告、总结报告、论文、著作,包括民族志,都是民族志工作的有形产品。民族志学者可以和参与者共享这些书面作品,以核实它们

的准确性；也可以和同事分享，以得到他们的评论和意见。

田野笔记就是民族志大厦的砖块和灰浆。这些笔记主要是由每天的访谈数据和日常观察组成的。民族志学者在田野工作的各个不同阶段中都会对研究作总结。民族志研究的最后一个阶段就是写作总结报告、论文或著作。民族志的特点通常是深描和逐字逐句地引用。互联网出版给我们提供了纸质出版物之外的另一种可供选择的出版形式。比起传统媒体，论文可以在互联网上更快得到发表，更快得到评论。

## 伦理

在民族志工作的各个阶段都弥漫着伦理问题。民族志学者要开展工作必须获得正式或非正式的知情同意【Informed Consent】。拍照片、录音或摄像都需要得到参与者的许可（书面作品如果要用于教育或商业领域也同样如此）。民族志学者应将他们的任务直言相告，解释他们计划研究什么，打算如何来研究。在需要的时候，他们必须作出保密的承诺，使用假名来保护被研究者的个人身份。民族志学者要完成工作需要得到共事者的信任。在民族志的从头到尾的每个步骤中，从一开始选择和定义某种文化及其问题开始，直到最后以写作结束这一经历，伦理都在提供指导。民族志学者在整个研究的生命周期中都站在伦理的十字路口，包括问题的启动和定义，出生；提交资助计划，童年；田野准备代表的是青少年时期和成年时期，最能体现田野工作的特点；而退休就是离开田野的时候。每个阶段都给民族志学者提出新的挑战，他要用伦理的考量来平衡各种方法论的关注点。这种民族志生命的事实使感觉更敏锐，最终会使这番努力的质量得到改进和提升（更多的细节请参见 Fetterman，1998）；美国人类学会（American Anthropological Association）的《伦理准则》（*Code of Ethics*）和《职业道德守则》（*Principles of Professional Responsibility*）可访问相关网站。

——David M. Fetterman

（杨可译　沈崇麟校）

## 参考文献


Bellman, B. L., & Jules-Rosette, B. (1977). *A paradigm for looking: Cross-cultural research with visual media*. Norwood, NJ: Ablex.

Birdwhistell, R. L. (1970). *Kinesics and context: Essays on body motion communication*. Philadelphia: University of Pennsylvania Press.

Fetterman, D. M. (1998). *Ethnography: Step by step* (2nd ed.). Thousand Oaks, CA: Sage.

Fetterman, D. M. (2002). Web surveys to digital movies: Technological tools of the trade. *Educational Researcher*, *31*(6), 29-37.

Fink, A., & Kosecoff, J. (1988). *How to conduct surveys: A step-by-step guide*. Newbury Park, CA: Sage.

Fowler, F. J. (1988). *Survey research methods* (Rev. ed.) Newbury Park, CA: Sage.

Geertz, C. (1975). *The interpretation of cultures*. New York: Basic Books.

Hall, E. T. (1974). *Handbook for proxemic research*. Washington, DC: Society for the Anthropology of Visual Communication.

Heider, K. G. (1976). *Ethnographic film*. Austin: University of Texas Press.

# 常人方法论(Ethnomethodology)

常人方法论既是学术研究领域的名称,也是该领域所研究的内容的名称。“常人方法论”这个词的字面意思是流行的或民间的方法。该术语由哈里德·加菲尔德(想要了解该术语起源的问题,请参阅 Garfinkel,1974)。加菲尔德(Garfinkel,1967)的著作《常人方法论中的研究》(*Studies in Ethnomethodology*)出版后,它成为社会学中一个确定的子域,尽管仍存在争议。常人方法论与社会学有着密切的联系,最近被美国社会学协会确定为一个正式的分支。我们说它与社会学的联系根深蒂固,原因是常人方法论对理解社会学的经典主题和问题提供了一种新奇的方法;在加菲尔德的研究(2002 年)的一卷新作中,他把它与涂尔干“社会事实”的基本概念相提并论。但是,常人方法论并不局限于社会学。连同谈话分析【Conversation Analysis】——作为一种研究社会互动的本构组织(constitutive organization)的派生项目,常人方法论已经成功进入人类学、社会语言学、计算机支持的协同工作、管理学研究和科学的社会研究等其他领域。

## 作为题材的方法

在任何一种明确的意义上,常人方法论本身就不是一种方法。正如民族音乐学不是音乐学的一个分支——是一种特殊的音乐技巧或传统——常人方法论也不是一种方法论。常人方法论的研究往往利用民族志程序,对连续活动进行“实时”的详细记录,但在许多情况下,学生负担最重的是要掌握和在程序上考虑,构成所研究的社会活动的专门方法。

常人方法论在其本源上具有社会学属性,因为它是对社会秩序产生的专业和非专业的方法的研究。因此,常人方法论学者要编写用于各种各样的日常和专门活动的方法的程序说明。这种方法的现代原型是对实验或观察程序的科学解释。正如萨克斯(Sacks,1992,pp.802-806)所指出的那样,科学方法的交流有赖于白描和处方。虽然科学的方法可能是高度技术性的,但它们也是更广泛的社会现象的具体实例,是一些能够让社区成员重现某种社区实践的指令。换句话说,方法和方法论解释都是规律的生活形式的反思性特征。萨克斯构想,常人方法论学者将会撰写方法论报告,但它不同于撰写那些实验和观察方案的职业方法论学者,常人方法论学者将深入那些丰富多样的普通方法中,正是通过这些方法,社会成员才得以协调和再现他们的活动。这些方法包括这样一些平凡的“成就”,例如,在红绿灯下过马路,调节对话的开启和结束,使会议有序进行,以及日常生活中许多其他的常规行动。

非专业和专业方法的常人方法论描述类似于科学家和其他专业人士书写的方法论。但是,与正式的方法论阐释不同的是,常人方法论描述包括计划、规则、算法、菜谱、地图、指南、格言、协议、谚语和其他正式指令,如何使用(以及废弃,使用不当或滥用)与情境相关的详细信息。这样的描述并不是反形式的,这里反形式的含义是拒绝相关的规则和公式。相反,常人方法论描述要揭示参与者在一个单一的行动过程中,部署规则、计划和其他有关方面的方法。常人方法论研究可能需要大量的训练和在特定工作环境中的参与观察【Participant Observation】。通常,常人方法论学者都会使用现场录制的

音频和视频，通过合作的、通常是即兴的、一连串话语和具体的实践来恢复随时形成的集体活动方式。通常，这种注重当时当地的现象学取向，排除了那些将情境话语降为标准代码和聚合模式的方法的使用。然而，常人方法论毕竟不是对主观意义的探寻。虽然它弥漫着存在主义现象学的气质，但常人方法论描述很少采取第一人称的经验反思形式[例外是萨德诺(Sudnow，1978，2001)阐明自己在发展演奏爵士钢琴方面的能力]。更多的时候，这样的描述都以第三人称进行，不享有"私人"或个体制高点的特权。这样的描述都有赖于对某种自然语言的精通及对一般实践的深刻理解。

## 作为题材的方法论

常人方法论的题材包括方法论的调查以及有条理的实践。换句话说，常人方法论既是方法论研究(在系统地反映了对方法和对方法的设计所做的努力这一意义上)，也是方法本身(按照一种或另一种方法论设计所开展的实践)。有一种观点认为，实践是在不知不觉中进行的，需要专业的分析师对其系统的组织进行阐述，常人方法论的处理方式则与这种观点形成了明显对比。对于常人方法论学者而言，对实践的阐释和实践本身是不可分割的，而让人感兴趣的正是这两者之间的交织。

加菲尔德的研究项目对所有的社会科学方法无一例外，他早期的一些研究深入探讨了社会科学研究实践，例如，对访谈答案编码和对经过组织的记录的分析。这些研究表明，社会世界的常识性知识中蕴含的特定判断，是构成数据节点和将实例分配给不同类别的关键。从 1970 年代开始，加菲尔德和他的一些学生把注意力转向自然科学。与其他常人方法论研究类似，实验室实践和高科技工作场所的调查，也关注那些司空见惯的细节，但为了显示这些平凡特征，必须要掌握专门的研究技术。这些研究的一个令人困惑之处在于，科学本来已经有了它们自己的、高度发展的方法论，何必还要多此一举。为了证明常人方法论研究不是庞大的方法论专业文献的一部分，常人方法论学者非常有必要去深入探究实践中那些在正式的方法论规范和报告中没有提到的、比较模糊，有时存在争议的特征。虽然社会建构路径与此有一定相似之处，但常人方法论学者并不表明或暗示，科学证据的这种特定生产方式会削弱这些证据的客观性。相反，对于常人方法论学者来说，科学生产的隐性特征是不可避免的；这些特征是客观性的来源，如同正是它的障碍。

——Michael Lynch

(郑晓娟译　沈崇麟校)

## 参考文献


Garfinkel, H. (1967). *Studies in ethnomethodology.* Englewood Cliffs, NJ: Prentice Hall.

Garfinkel, H. (1974). On the origins of the term "ethnomethodology." In R. Turner (Ed.), *Ethnomethodology* (pp. 15-18). Harmondsworth, UK: Penguin.

Garfinkel, H. (2002). *Ethnomethodology's program: Working out Durkheim's aphorism.* Blue Ridge Summit, PA: Rowman and Littlefield.

Sacks, H. (1992). *Lectures on conversation* (Vol. 1). Oxford, UK: Blackwell.

Sudnow, D. (1978). *Ways of the hand.* Cambridge, MA: Harvard University Press.

Sudnow, D. (2001). *Ways of the hand* (Rev. ed.). Cambridge, MA: MIT Press.

## 民族志统计学(Ethnostatistics)

民族志统计学使用了常人方法论【Ethnomethodology】的概念来研究社会科学家在社会研究过程中创建的统计数字的生产、解释和展现的意义制造的实践。民族志统计学关注社会语境下的量化推理和语言运用,以便理解专业工作者是如何制造出那些创造了一种共享社会现实感的数字。

民族志统计学与统计学【Statistics】不同,统计学是使用一些技术和规则控制的程序来对研究数据进行计数、汇总和估计。统计学领域不研究制造数据所需的实用知识或不同情境如何影响科学家的行为。因此,正式的统计程序和规则无法对制造统计数字提供完备的指导。民族志统计学则关注描述和理解进行实际统计所需的技术知识和程序。

民族志统计学有三个层次。第一层次的民族志统计学使用民族志方法来对统计数字的社会生产过程进行浓描【Thick Description】。例如,古布里乌姆和布克霍尔特(Gubrium & Buckholdt,1979)研究了医学项目有效性的数量量度的专业制造过程,并证明技术规则不足以解释计数实践和统计数字。专业人士用来确定哪些行动是应该在计数的标准中,除了技术维度的标准之外,非正规的标准,例如病人依从项目目标的动机也应在考虑之列。

第二层次的民族志统计学对实际使用的统计量进行研究,将含义的计算机模拟方法【Computer Simulation】用于解释那些有违反各种技术假定之嫌的统计量。一项研究检验了社会科学中一个常见假定的效度,即在统计分析中,定序数据可以通过将连续的数字(例如1—7)分配给数据类别,来转化为定距数据。这种做法假定数据转化对结果和发现的解释没有影响。民族志统计学的研究却证明这种数据转化将会产生"本质上偏离真实数据"的测量值和统计值(Gephart,1988,p.38)。假设检验和路径分析【Path Analysis】的数据解释也会受到影响。因此,第二层次的民族志统计学揭示了常见的测量方法可能会造成对结果的曲解。

第三层次的民族志统计学使用修辞学【Rhetoric】和文本分析方法来效验文件和文本中统计数字展示的那些有说服力的特性。一项关于定量方法论论文是如何创造意义和解释数字的研究发现,意义是由用以描述和解释数字的形容词和统计描述所嵌入的情境所共同创造的。一个数字的绝对值并不决定对这个数字作出的解释,且统计显著性检验的作用也十分微小。这样的情况的确是存在的:一位作者在一篇定量研究论文中称之为"相当大"的相关系数常常在0.1之内,而同样的相关系数,同一作者在同一篇文章的另一处则把它称为"比较小"。

通过审视数字的意义实际上是如何被创造的,民族志统计学将科学研究中的定量和定性方面整合在一起。通过将定量研究与社会生活中的定性基础相联系,将人文意义重新赋予定量研究(Gephart,1988)。

——Robert P. Gephart, Jr.

(郑晓娟译 沈崇麟校)

### 参考文献

Gephart, R.P.(1986).Deconstructing the defence for quantification in social science: A content analysis of journal articles on the parametric strategy. *Qualitative Sociology*, *9*, 126-144.

Gephart, R. P. (1988). *Ethnostatistics: Qualitative foundations for quantitative research*. Thousand

Oaks, CA.: Sage.

Gephart, R. P. (1997). Hazardous measures: An interpretive textual analysis of quantitative sensemaking during crises. *Journal of Organizational Behavior*, *18*, 583-622.

Gubrium, J.F., & Buckholdt, D.R.(1979).Production of hard data in human service organizations. *Pacific Sociological Review*, *22*, 115-136.

## 行为发生学(Ethogenics)

行为发生学这一词指的是1970年代英国牛津所创立的一个新的社会心理学研究范式(Harré & Secord,1972),它的出现是对美国主流心理学研究方法和研究理论输入的有意识反应。行为发生学受到了常人方法论【Ethnomethodology】(Garfinkel,1967)、戈夫曼的微观社会学和社会文化人类学的影响。

行为发生学拒绝承认那种关于抽象层次的普遍性预设,认为这样的抽象抹杀了具体的社会现象。他们拒绝将个体作为应激/反应模式中的被动点来对待,而是将个体看成是在参与项目时,依据当地的规则和传统与他人进行互动的主动点。他们特别指出了对统计方法的错用,尤其是那些常见但却仍然常被误用的,从统计分布中得出有关个人特质的结论。

行为发生学强调将社会互动过程中的真实片段,作为依照当地关于恰当行为的规则、传统和习俗的指示而有序展开的意义结构来分析。关于社会行动的动力学研究,例如,朋友的结交,代替了有关友谊这种静止状态的条件的研究。对当地习俗和传统的关注导致了对社会人类学的兴趣,并将其作为社会心理学中的一种新研究范式的关键要素。

这种新的研究范式中的方法论特征,要求使用行为/行动特征来识别社会现象的意义,以及这些意义如何被参与者所理解。对真实生活片段的观察代替了实验【Experiment】。这种方法使用于对参与者的辩护和解释性说明的分析,来识别研究项目中社会互动所实现的成果,以及社会互动所遵循的规则和传统。

这一新研究范式的研究结果是一系列特定情境下的意义和对情境敏感的规则,这些意义和规则能够解释发生的社会现象片段的模式,并将社会现象片段看成一个有意义的社会行动的实际序列。这导致了它与新兴的话语心理学(discursive psychology)相结合,后者指向类似的最终产品领域。"规则"的概念代替了一系列规范性限制,我们认为它对于形塑行动流是非常有效的。

行为发生学非常强调描述性和解释性模型的建构。这些模型中最为典型的是戏剧法模型。社会现象的片段可以被看成类似于舞台表演,有布景、演员以及相互影响的行动。通过将社会现象片段与游戏和典礼进行类比,这个基本的模型得到了进一步发展和细化。

这一观点在发展社会心理学中的进一步应用,与(社会)建构主义【Constructionism】和文化心理学的发展形成了紧密的同盟关系(Cole,1996)。

——Rom Harré

(郑晓娟译 沈崇麟校)

### 参考文献

Cole, M. (1996). *Cultural psychology: A once and future discipline*. Cambridge, MA: Belknap Press of Harvard University Press.

Garfinkel, H. (1967). *Studies in ethnomethodology*. Englewood Cliffs, NJ: Prentice Hall.
Goffman, E. (1959). *The presentation of self in everyday life*. New York: Doubleday.
Harré, R., & Secord, P.F. (1972). *The explanation of social behaviour*. Oxford, UK: Blackwell.

## 动物行为学(Ethology)

动物行为学是一个将发展于非人类的动物研究中的模型和方法,应用于人类研究学科中的好范例。作为查尔斯·达尔文进化论观点的忠实信徒,动物行为学通过它对行为的多层次理解原则的恪守,为进化论的理论改进作出了最有价值的贡献,这些层次包括:(a)最近因果关系(在起源上可能是生理性或环境性的);(b)终极因果关系(在强化生存和/或繁衍适应性方面行为的适应性功能);(c)种系发生(相近物种中的行为模式);(d)个体发生(行为的发展历程)(Tinbergen,1963)。这个大概的研究框架让动物研究者离开他们可以控制的实验室环境,去寻找他们感兴趣的物种已经适应的自然栖息地,在那里对它们进行观察。例如,洛伦兹通过观察小鹅一从鹅蛋中孵化出来就开始跟随自己的母亲或者洛伦兹本人的现象,发现了鹅群中的留下印记行为(Lorenz,1981)。只有在完成对这一观察工作的描述阶段后,洛伦兹才开始设计有关留下印记的实验,例如,引入孵化出来的时间点作为一个变量,但是仍然将实验置于德国湖边的自然环境中。因此,动物行为学经常被认为是有关在动物的自然环境中的动物行为的研究。

描述自然环境中发生的行为,并不意味着放弃精密研究的科学观点。一个动物行为学的学者会花费大量时间为所研究物种的行为进行分类编目,并为该物种创造"动物行为详述"。需要讨论的典型问题包括哪些行为看上去是共同发生的?在自然环境的什么条件下发生?这些行为是否直接指向特定目标?如果目标没有实现,动物们会怎么做?这些行为是否组成一个固定的行动模式?这些行为出现的频率怎样?这些行动之间是否存在个体差异?这些差异是否与性别、年龄或社会地位有关?行动的差异是否与一天中的时点、季节、温度或其他环境因素有关?通过这个过程,动物行为学的学者最终会回到前面已经介绍过的廷伯根的四层次分析来进行理论的修正和完善。如果有需要,动物行为学学者也会采用实验方法来解答某些问题。例如,当某个特定的环境刺激被假设为是某个行为的"释放因素",动物行为学学者可能会进行一个剥离实验,即消除该环境刺激因素,来看感兴趣的行为是如何改变的。通过控制刺激因素释放的时间,动物行为学学者能够辨识出这些行为发展的敏感时期(之前称为关键时期),例如,鹅群中的留下印记时期和狗中间的社会化时期。因此,动物行为学研究可以被认为是以一种在定性和定量路径之间游动的系统性方法进行的行为研究。

在1960年代和1970年代,一些受训于传统动物行为学领域的科学家进一步开始研究人类,这方面我们可以从艾伯-埃艾斯费尔特(Eibl-Eibesfeldt,1989)的著作中看出,他与洛伦兹共同研究并记录了世界各地的人类情感沟通。同样是在这一时期,约翰·鲍尔比和玛丽·安斯沃思受到达尔文的进化论观念和哈利·哈洛的灵长类动物方法论影响,他们有关"母亲-婴儿"依恋关系的研究得到了心理学领域的广泛关注。但是,很多欣赏这一方法论的学者却并不赞

同鲍尔比的进化论观点，尤其是他认为依恋关系的终极目标就是婴儿的生存。

动物行为学对观察研究【Observational Research】使用了非常严谨的路径规范。许多资料中详细记述了动物行为学研究中方方面面的标准，包括问题的概念化、对被研究者的抽样【Sampling】、时间和空间的记录、事件的记录、对环境刺激物的测量和对观察者本身信度的评估（参见，例如 Blurton Jones，1972；LaFreniere & Charlesworth，1983；McGrew，1972）。例如，如果一位学者对于学龄前儿童参与挑衅性互动行为的频率感兴趣，他需要首先定义挑衅行为，并将符合这一定义的具体行为进行列示（推、咬、踢等）。接下来，在研究的初始阶段，研究者可能很容易记录可观察到的挑衅行为（随机抽样）；或者研究者也能比较容易地记录每分钟能够观察到的这些行为（事件抽样）；或者研究者可以记录每 15 秒某个挑衅行为是否发生（一零抽样）；或者研究者可以观察某个特定的儿童，记录他在一段给定的时间内是否展示出挑衅行为（聚焦抽样）；或者观察者可以记录在挑衅行为前后所发生的行为（顺序抽样）。在对一家托儿所学龄儿童进行的动物行为学研究中，通过使用这些方法，麦格鲁（McGrew，1972）指出，（除了其他发现外）挑衅与带有输赢性质的打斗有关，并且老师的评价与科学研究中设定的观察行为之间有高度相关关系。

有时，为了进行更加细致严密的分析，需要使用慢镜头下的影片或视频分析，或实验室中的录音分析，或者使用破坏嗅觉刺激的化学物质分析。在人类研究中，面部肌肉运动被分解为不同的单元，在卡尔·格拉默、芬克和伦宁格（Grammer，Fink & Renninger，2002）的研究著作中，姿势和运动质量的研究是通过计算机建模来实现的。

在评价动物行为学研究的重要性时，廷伯根（Tinbergen，1963）曾指出，“许多人认为，在心理学匆匆踏入 20 世纪成为一门受人尊敬的科学的过程中，它略过了其他自然科学必经的初始描述阶段，因此很快就与自然现象失去了联系。”（p.411）动物行为学，由于同时提供了描述性和实验性研究方法的优势，能够在总体上为社会科学提供尚未被发现的启发价值。除此之外，动物行为学的进化论基础提供了一个非常有包容性的方法，为我们提供了一条走出无益的自然/教化的二元对立的出路。

为了表彰康拉德·洛伦兹、廷伯根和卡尔·弗里希几位学者在动物行为学方面的杰出工作，他们被授予了 1973 年的诺贝尔生理学或医学奖。

——Carol Cronin Weisfeld

（郑晓娟译　沈崇麟校）

## 参考文献

Blurton Jones，N.（Ed.）.（1972）.*Ethological studies of child behaviour*. London：Cambridge University Press.

Eibl-Eibesfeldt，I.（1989）. *Human ethology*. Hawthorne，NY：Aldine de Gruyter.

Grammer，K.，Fink，B.，& Renninger，L.（2002）. Dynamic systems and inferential information processing in human communication. *Neuroendocrinology Letters*，*23*（Suppl.4），15-22.

LaFreniere，P.，& Charlesworth，W. R.（1983）. Dominance，attention，and affiliation in a preschool group：A nine-month longitudinal study. *Ethology and Sociobiology*，*4*（2），55-67.

Lorenz，K.（1981）. *The foundations of ethology*. New York：Simon and Schuster.

McGrew，W.（1972）. *An ethological study of children's behavior*. New York：Academic Press.

Tinbergen，N.（1963）. On aims and methods of ethology. *Zeitschrift für Tierpsychologie*，*20*，410-433.

## 评估焦虑(Evaluation Apprehension)

评估焦虑是反应效应的一种。已经证明它会出现在实验中。实验的应试者会因为自己将要被实验者测试而忧虑,使测试变得很差。所以在实验研究中,有时我们也会推荐使用那种能使实验参与者明白实验的目的并非是要对他们的表现进行评价的幌子。

——Alan Bryman

(沈崇麟译校)

## 评估研究(Evaluation Research)

评估研究是一种系统的以数据为依据的探究方法,其目的在于确定一个项目、产品、组织、干预或实施的整改措施的成效或价值。评估研究应用社会科学和相关的探究方法,系统地收集有关活动、特征和整改措施结果的信息,使我们能有充分的根据对目标的实现、整改项目的效益、成本和收益的确定和/或作出进一步的决定等问题作出判断。

数以亿计的美元花费在了解决贫困、疾病、愚昧、失业、精神创伤和不平等这些问题上,为了改进人们的境况而采取的许多措施不一而足。那些向这些社会病宣战的项目应该如何评价呢?如何从那些无效果的项目辨别出效益来呢?数据应怎样来收集,并以提高项目的效益和对决策提供支撑的方式报告呢?这些问题正是评估研究的领域。

### 评估和实验社会

爱德华·萨奇曼(Suchman,1967)用他锐利的观察在他有关评估研究的富有开创性的著作的开篇中写道:“我们这个世纪最吸引人的思想之一是科学可以为我们提出解决社会问题的办法这一概念。”(p.1)空想家,如唐纳德·坎贝尔(Campbell,1991)将评估概念化为一种新社会:实验社会(experimenting society)的中心内容。他于1971年在美国心理学协会(American Psychological Association)宣布了这一思想(Campbell, 1991):

“实验社会将是一种设法为各种经常发生的问题提出解决方案和对其实施结果进行冷静和多维度的评估,并在评估显示改革无效或有害的时候转向另一种备择方案的社会。不过,直至今日,我们还没有这样的社会。”(p. 223)

### 历史背景

在美国,评估的正式专业根源可追溯到1960年代由(当时的美国总统约翰逊提出的社会福利方案)大社会(Great Society)引起的立法。这一方案的目的在于消灭贫困。大规模的联邦卫生项目的建立,包括社区精神健康中心的建立,必然伴随对评估的需求,而评估的费用常常占项目预算的1%~3%。其他的重大项目则在住房、就业、服务整合、社区规划、城市改造和福利事业等方面建立。社会行动项目的迅速成长创造了对政府项目进行系统的经验评估的需求。于1965年通过的“美国初等和中等教育法案”对更为全面的评估方法的产生有着巨大

的贡献。大量联邦资金流向废止种族歧视、创新、补救教育、更大的机会平等、教师培训和更高的学生学业成就等项目。与之相伴的是对评判各种各样有关国家儿童项目的数据的巨大需要。

项目评估是专业工作中的一个独特领域,在大规模社会实验和政府干预期间有过两个教训。第一个是意识到没有足够的钱去做需要做的所有事情;第二个是即使有了足够的钱,但解决复杂的人和社会问题需要比钱更多的东西。因为并非每一件事情都是可为的,必定会存在一种根据,可据此确定什么事情是值得做的。评估研究通过提供载有对效益进行判断的可靠的数据承担起为决策提供依据的任务。

然而,早期对评估的愿景都集中在了它对资金的分配决策的引导和区分联邦项目优劣的作用上了。随着评估的实施,一种新的作用出现了:在项目的实施过程中帮助改进项目。人们要求评估者不只是对项目的整个效应作出最终的判断,也要去收集过程数据并给出反馈,以在项目实施过程中帮助解决设计规划问题。

## 评估研究的主要发展

### 1.评估,既是一门学科也是一种专业

评估研究开始作为应用于特定问题的方法,它都在那些已经建立起来的学科,如经济学、心理学和教育学之内。当评估研究者来到一起分享他们的经验和观点时,评估便变成了专业实践的一个独特的领域。定期举行年会的正式的专业协会和专业杂志则在1970年代和1980年代形成,其中有全美评估协会(the American Evaluation Association)、加拿大评估学会(the Canadian Evaluation Society)、澳大利亚评估学会(Australasian Evaluation Society)。

评估也变成了一个调查的学术领域,专门从事确定项目的得失、价值、重要性和效益,从而走出了其他学科的阴影,使自身成为一门学科,甚至如哲学家和评估家米歇尔·斯里文(Scriven,2001)所言,它已经成为一门超学科(transdiscipline)。

“评估是超学科的精英团体之一,在这一背景中使用的一个学术名词代表了那些为其他学科服务而著名的学科,尽管它们也有自己的自治地位。这些超学科涵盖的领域广泛,从小的但重要的一些学科(如测量),到一些大多数定量学科都不可或缺的重大学科(如概率和统计学),包罗万象,甚至也包括逻辑学和评估。”(p.304)

### 2.完备的评估标准

评估的专业化使评估自身的标准呼之欲出。在专业确认和采用自己的标准之前,判断评估的准则几乎与传统社会科学中的判断研究准则没有什么区别,注重技术的质量和方法论上的严谨。经过17位由12个专业组织任命的委员会成员5年时间的艰苦工作,加上几百位从事评估研究的专业人士的援手,标准于1981年问世,而后,于1994年作了修订(Joint Committee on Standards for Educational Evaluation,1994)。该标准引人注目地表明评估研究已经在实践中成熟了,它要求用四组准则来判断评估:实用性、可行性、正当性和精确性。

### 3.日益引人注意的评估应用

在1960年代末,我们清楚地看到,大社会项目的评估被极大地忽略或政治化了。韦斯(Weiss,1972)第一个提请人们注意那些尚未被充分利用的问题,他把这一问题看成是评估研究中最为重要的问题之一,他在对项目作早期评审时说道:“评估结果并没有对项目的决策发挥显著影响。”(pp.10-11)

这也不是日益增长的使用所面临的唯一的挑战。评估者越来越担心评估研究的发现被误用。马尔温·阿尔金(Alkin,

1990）一位早期的用户取向的评估理论家，强调评估者必须致力于使用的合理性（appropriate），而不要只是注意使用的数量。

因此，评估面临双重的挑战：必须在加强合理使用的同时，消除不合理的使用。在对这些问题作出回应时，一种称为注重使用【Utilization-Focused】的评估方法（Patton，1997）应运而生。这种方法强调实施评估要达到预期的使用者预期的使用目的，注重使用的评估。在设计阶段就注意从实施开始到结束的每一个细节，考虑每一环节将对使用有什么影响。一种心理学的使用加强了注重使用的评估：如果预期的使用者了解和感觉到评估过程和发现的所有权，那么他们就更有可能使用评估；如果他们曾经积极地参与了评估，那么他们就更有可能了解和感觉到所有权；通过让主要的预期的用户积极参与评估过程，评估员在使用中培训用户，准备好使用的基础，并在评估过程中的每一步加强预期的评估使用。

注重使用的评估也注意过程的使用（Patton，1997），与发现的使用形成对照。过程的使用指的是思想和行为的变化，以及程序和文化中的项目或组织的变化，它们发生在涉及那些在评估过程中发生的作为学习结果的评估。过程的使用如项目工作人员在一个评估之后评估的一样："对于我们的项目的影响，来自发现的结果并不多，而是来自评估要求的整个思考过程。"

### 4.可贵的方法论多样性

在评估研究发展的早期，爆发过有关定量/实验方法与定性/自然主义方法孰优孰劣的争论。这一争论爆发出思想的火花（Patton，1997）。在评估者需要知道并使用各种方法，以对特定的评估问题存在的细微差别以及特定利益关系人对需求的个人喜好作出反应时，形成专业的共识。现在评估研究设计的焦点已是特定目的和预期使用的方法的恰当性，不再坚持某些是绝对正统的（观点），认为只有定量或定性方法才一定是内在的首选的方法。联合委员会的标准（The Joint Committee Standards，1994）给两种方法以相同的权重。在这个领域中，人们已经认识到，只要有可能，就应该使用多种方法——包括定量和定性，只有这样的方法才是最有价值的，因为每一种方法都有自己的优势，且一种方法还常常可以克服另一种方法的弱点。

### 5.许多不同类型的评估和新评估者的作用

在方法上变得越来越多样化的同时，由于许多新的方法、模型和评估者的作用的出现，评估研究的实践变得越来越多样化。最初，评估只不过是对一个项目的目标实现进行评价，对功过和价值作出整体判断，这样的评估称为总结性评估（summative evaluation）或判断取向评估（judgment-oriented evaluation）。总结性评估测量结果和影响，以及项目活动和构成项目模型或理论的结果之间的因果联系。逐渐地，人们开始把更多的注意力放在了通过收集特定的反馈与识别优点和缺点，以用来对项目作改进，这样的评估常常称为形成性评估（formative evaluation）或改进取向的评估。形成性评估考察实施情况、项目过程和项目对客户和条件变化的适应性。虽然有时形成性评估用来为总结性评估作准备，但越来越多的是用作改进取向的评估，它已成为为持续地改进质量而做的不懈努力的一部分。

判断取向和改进取向的评估二者都涉及结果的工具性（instrumental）使用，这意味着紧随评估之后的将是一个决定或行动。另一方面，发现的概念的使用增加了我们的知识，这样评估的发现就会影响论题、观点或政策选择。产生的知识可以具体到明确一个项目的模型、检验理论、区别干预的类

型、明确如何测量结果、得到经验教训和/或详细政策的选择。近年来,产生知识的评估已经变得非常重要。

另一种转变涉及外部和内部评估。由于早期的评估都注重总结性评估,因此大多数评估都委托外部机构进行,以确保独立性和可信性。随着评估越来越多地强调持续的质量改进,内部评估开始兴旺起来。

为了解决特别的问题或达到特定的目的,一些新的评估方法应运而生。无目标评估(goal-free evaluation)给我们提供了一种替代基于目标的评估方法。这种方法旨在发现客户的主要需求是否得到满足的同时,还要捕捉副作用和意料之外的影响。反应性评估(responsive evaluation)是一种捕捉和顾及差别与利益相关者观点冲突的方法。强能评估(empowerment evaluation)是一种帮助参与者诉说自己的故事,从而提高他们从事评估过程的能力的方法。包容性评估(inclusive evaluation)强调为了包容弱势和享有较少特权的群体的观点必须作出的特别的努力。理论驱动型评估(theory-driven evaluation)在真的项目设置中正式检验社会科学理论。实在论评估(realist evaluation)用实在论【realism】哲学构建评估框架。民主评估(democratic evaluation)注重用评估支持对话和商议以提高社会公正。在发展性评估(developmental evaluation)中,评估者是一个正在持续进行的改进过程中的、合作进行概念化、设计和检验新方法的团队中的一员。

这些方法只是给我们提供了评估研究中出现的丰富多彩的选项菜单的一个样本。

### 6.需要的新技巧和新素质

评估研究的新方法拓展了成为一个优秀的评估者所需的技巧。评估研究的发展已经超出了简单地应用社会科学的方法来评价一个项目的目标实现的范围。为了确保评估的功用,评估者需要交流、谈判、解决冲突和团队的组织引导方面的技巧。为了满足可行性标准,评估者需要富有政治经验、求真务实和创造性,还要能掌控评估过程,以能按预定计划和在预算之内产生评估的结果。为了满足合理性标准,评估者需要能跨学科地运用各种各样的方法和工作。这些新的需求正在改变评估研究者的培训方式。

### 7.技术和评估研究

在现代社会中,技术的主导作用已经对评估研究产生了很大影响,就像评估研究对其他的事业已经产生了影响一样。互联网打开了交流的新渠道,它通过网络和邮件用户清单服务将全世界的评估者联系在一起。软件的发展,尤其是用于定性数据的软件的发展大大推动了数据分析的发展(Patton,2002)。计算机图形和文字处理软件的发展,改变了评估报告的撰写和散播方式。将评估报告放在万维网(World Wide Web)上开阔了人们得到评估发现的途径,推动了基于实践的证据的散播,使我们能通过不同的评估更多地总结经验教训,它也成了一种知识产生的方式。用互联网手机数据的新技术(如在互联网上进行的调查)改变了数据收集的方式。录像和摄影被用于数据的收集和报告的撰写。管理信息系统辅助我们进行能汇总用于项目评估目的的常规监测数据的收集。技术的发展正在以各种方式影响着评估研究。

### 8.全世界对评估的需求

第一届国际评估会议于1995年在温哥华举行,它由当时很活跃的三大评估协会,加拿大、澳大利亚和美国的评估协会组织发起。与会者有1 500人之多,来自61个国家。这次会议有着重大的历史意义。在21世纪初,有40多个国家已经建立了评估协会,除此之外,还有欧洲评估学会(the European Evaluation Society)、非洲评估协会

(the African Evaluation Association)和国际评估协会(International Evaluation Association)等国际和地区协会。国际机构也开始用评估来全方位地评价在发展中国家进行的各种开发工作。全球的互动决定评估研究的未来,给这个专业注入了新的能量,全世界的评估工作者相互学习,一起为发展适合不同文化、社会和政治背景的评估研究方法而努力奋斗。

——Michael Quinn Patton

(沈崇麟译校)

## 参考文献


Alkin, M.(1990). *Debates on evaluation*. Newbury Park, CA: Sage.

Campbell, D. T.(1991). Methods for the experimenting society. *Evaluation Practice*, *12*, 223-260.

Joint Committee on Standards for Educational Evaluation. (1994). *The program evaluation standards*. Thousand Oaks, CA: Sage.

Patton, M. Q.(1997). *Utilization-focused evaluation: The new century text*(3rd ed.). Thousand Oaks, CA: Sage.

Patton, M. Q. (2002). *Qualitative research and evaluation methods*(3rd ed.). Thousand Oaks, CA: Sage.

Rossi, P. H., Freeman, H. E., & Lipsey, M. (1998). *Evaluation: A systematic approach* (6th ed.). Thousand Oaks, CA: Sage.

Scriven, M.(1991). *Evaluation thesaurus*(4th ed.). Newbury Park, CA: Sage.

Scriven, M. (2001). Evaluation: Future tense. *The American Journal of Evaluation*, *22*, 301-307.

Shadish, W. R., Jr., Cook, T. D., & Leviton, L. C. (1991). *Foundations of program evaluation: Theories of practice*. Newbury Park, CA: Sage.

Shadish, W. R., Jr., Newman, D. L., Scheirer, M. A., & Wye, C. (1995). *Guiding principles for evaluators*(New Directions for Program Evaluation, No.66). San Francisco: Jossey-Bass.

Suchman, E. A. (1967). *Evaluative research: Principles and practice in public service and social action programs*. New York: Russell Sage.

Weiss, C. H.(1972). A treeful of owls. In C. H. Weiss(Ed.), *Evaluating action programs* (pp. 3-27). Boston: Allyn & Bacon.

Weiss, C. H., & Bucuvalas, M.(1980). Truth test and utility test: Decision makers' frame of reference for social science research, *American Sociological Review*, *45*, 302-313.


# 事件计数模型(Event Count Models)

在许多研究背景中,因变量是作为计数来量度的。一个计数变量是一个代表观察的事件(或人、项目等)数的非负整数。其目的是用合并的预测变量来揭示观察的计数中的变差。普通最小二乘回归【Ordinary Least Squares Regression】分析,要假定有一个连续的因变量,因而对计数数据是不合适的。

计数模型最好放在广义线性模型【Generalized Linear Model】的背景下来考虑。任何只允许非负整数的概率【Probability】分布都可以用来描述这种因变量。对于计数模型的精细研究可架起一座通向广义线性模型的非常好的桥梁。

假设每一个观察 $i$ 有一个 $m$ 个预测变量的向量:$X_i=(x_{1i},x_{2i},\cdots,x_{mi})$。假设估计的系数 $b=(b_1,b_2,\cdots,b_m)$,可以被用来预测因变量 $y_i$ 的期望值。我们的第一直觉可能会告诉我们用诸如下面这样的线性

模型

$$E(y_i \mid X_i) = X_i b = \sum_{j=1}^{m} b_j x_{ji}$$

这种方法的缺点在于 $y_i$ 的期望值有可能是负的，这就违反了所有观察的计数为 0 或大于 0 的基本假定【Assumption】。为了避免这样的问题，一般都要使用数学变换【Transformation】。在大多数情况下，我们都使用指数变换：

$$E(y_i \mid X_i) = \exp(X_i b) = e^{X_i b} = e^{\sum_j b_j x_{ji}}$$

最大似然估计法【Maximum Likelihood Estimation】被用于导出 $b$ 的估计值。

最大似然的计算需要对 $y_i$ 的特定分布作出假定。在 1990 年之前，使用最为普遍的分布是泊松分布【Poisson Distribution】。给定输入变量 $X_i$ 和参数估计值 $\hat{b}$，那么恰好观察到事件 $y_i$ 的概率是：

$$\Pr(y_i \mid X_i) = \frac{\exp(-e^{X_i b})(e^{X_i b})^{y_i}}{y_i!}$$

参数 $b$ 的估计值是有选择的，以使观察的样本最有可能。

泊松分布比较简单且切实可行。如果 $E(y_i \mid X_i)$ 比较小，那么泊松分布会大幅倾斜，有非常小的模态值和向右的长尾。随着 $E(y_i \mid X_i)$ 的增大，分布的形状也随之变化，在外表上变得相当正态（大值泊松分布和正态分布之间的相似性有时被用作证明普通最小二乘回归用于计数数据的正当性）（有关泊松模型的调查和社会科学的应用请参见 King，1988）。

泊松分布有一个不太好的性质：它的方差【Variance】等于均值。在它与观察数据拟合时，我们常常会发现过度散布问题；观察数据的方差大于基于泊松分布的期望（方差）。

许多学者建议用负二项分布【Negative Binomial (NB) Distribution】来取代泊松分布。NB 模型有与泊松分布相同的均值，但它有较大的方差。它还有另一个有用的性质。我们可以用那种允许以可加的干扰（误差）项【Disturbance (or error) Term】的形式的个体异质性【Heterogeneity】的方式来书写原模型。令 $u_i$ 代表（未观察到的）个体变差：

$$\begin{aligned} E(y_i \mid X_i) &= \exp(X_i b + u_i) \\ &= e^{X_i b + u_i} \\ &= e^{\sum_j b_j x_{ji} + u_i} \\ &= e^{\sum_j b_j x_{ji}} \times e^{u_i} \end{aligned}$$

如果 $u_i$ 刚好等于 0，那么这个模型就降格为泊松模型（$e^0 = 1$）。然而，如果我们假设 $u_i$ 有 Gamma 分布，那么观察的 $y_i$ 分布将会有负二项分布。我们一般假定 $E(u_i) = 0$，这种个案之间观察的“平均”异质性就没有影响。这就是说，泊松或负二项模型的期望值是相同的。然而，在方差 $u_i > 0$ 时，负二项分布的离差更大。朗（Long，1997）的教科书中，对这一问题的处理方法作了介绍。

我们可以得到计数模型框架的很多扩展，在许多高级统计学文献中有很多有关的推介性文章。这种路数可扩展到那些计数中观察到的零计数的数目大于期望的场合。卡麦伦和特里维迪（Cameron & Trivedi，1998）的著作中有有关计数模型问题的广博介绍。

——Paul E. Johnson

（沈崇麟译校）

## 参考文献

Cameron，A.C.，& Trivedi，P. K.（1998）. *Regression analysis of count data*. Cambridge，UK：Cambridge University Press.

King，G.（1988）. Statistical models for political science event counts：Bias in conventional

procedures and evidence for the exponential Poisson regression model. *American Journal of Political Science*, *32*, 838-863.

Long, J. S. (1997). *Regression models for categorical and limited dependent variables*. Thousand Oaks, CA: Sage.

# 事件史分析(Event History Analysis)

事件史分析是一种使研究者不仅能确定一个事件是否发生,而且还能确定它是什么时候发生的技术。事件是从一种状态到另一状态的变化,因变量是到事件发生的时间。事件史分析是一种理想的适合纵向变化研究的方法,可看成 Logit/Probit 分析【Logit/Probit Analysis】的扩展和考虑事件的时机的事件计数模型【Event Count Model】。

纵向数据中出现的两种复杂的情况推动了事件史分析的使用。第一种情况是有关持续时间(一个观察在某一种特定状况花费的时间量)的信息不完整,数据中存在截删观察。例如,在研究结束之前观察并未经历感兴趣的事件,或因为在追踪调查时观察丢失了(丢失的原因可能是该调查对象搬出了调查地),这样的情况就可能会发生。第二种情况是时变性协变量(或自变量)有随时间变化的值。例如,在国会选举中的挑战者进入时间的研究中,现任议员能够募集的资金量在整个选举周期中可能是一个时变性解释变量。事件史技术可以马上就把截删的观察和时变性解释变量结合在一起。将时变性协变量包括进来可导致一个事件发生的风险是如何随协变量的值的变化而变化的新颖信息。

## 历史发展

事件史分析也称为持续时间分析、生存分析或可靠性分析(duration, survival or reliability analysis),具体名称随涉及的学科而异(后两个专用名词则分别涉及医学和工程学)。事件史分析的早期应用涉及卡普兰和迈耶(Kaplan & Meier, 1958)使用的生命表分析,但它的历史根源甚至可进一步上溯到 1600 年代后期(Hald, 1990)。在第二次世界大战期间,这一技术的使用有很大的增长,因为事涉军事装备的期望可靠性。事件史分析的使用的另一个扩张时期归功于考克斯(Cox, 1972)的开创性工作,正是他开发了半参数技术的结果。他的工作可能是开拓性的,可被视为 20 世纪最高统计学成就之一。由于约束较少的建立在计数过程数学上的半参数 Cox 回归模型和它的各种扩展的出现,事件史分析在医学和社会科学中的应用也有了很大的增长(Therneau & Grambsch, 2000)。

## 参数和半参数模型

参数和半参数模型二者都可供我们在事件史分析中使用。研究机械系统的分析师一般都使用参数模型,这种模型假定到达我们感兴趣的事件发生的时间服从某种特定的分布,如指数分布。人类行为和生物学研究一般都使用约束较少的半参数的 Cox 模型,这种模型未对持续时间的特定分布形式加以设定。布洛斯费尔德和罗韦尔(Blossfeld & Rohwer, 2002, pp. 180, 263)指出,社会科学理论很少能对某种特定的参数分布提出正当的理由,因而都提倡使用 Cox 模型。

事件史模型的估计和解释的一个关键概念是风险率【Hazard Rate】。风险率给出

了那些假定在到时间 $t-1$ 一直存活的观察，到时间 $t$ 会衰亡（或持续时间结束）的比率。换言之，它可被解释为一个特定的观察在特定时间一个事件将会发生的概率。风险组则指那些仍处在经历感兴趣的事件的"风险"的观察。一旦观察在时间 $t$ 经历了事件，观察便退出了风险组，并不再是在时间 $t+1$ 分析的数据的一部分。在感兴趣的事件取决于它的历史时，风险率便会有很大的吸引力。例如，如果一场战争已经持续了 $t$ 时期，那么它将在后续期结束的可能性有多大呢？Cox 模型的风险率可写为：

$$h(t \mid X_i) = h_0(t) e^{X_i \beta}$$

式中 $h_0(t)$ 是一个（未设定的）基线风险函数，而 $X_i$ 则是观察 $i$ 的协变量。

## 模型的假定和模型的拟合

在拟合事件史模型时，一个重要的要检查的假定是比例风险假定。大多数事件史模型，包括 Cox 模型，都假定任何两个在一个或多个协变量上有着不同的值的风险函数的不同，都只是比例因子不同而已。换句话说，基线风险率是随时间，而不是跨个体变化的，这样，个体 $i$ 和 $j$ 的风险的比率对时间 $t$ 是独立的，因而对所有的 $t$ 都是不变的：

$$\frac{h_i(t)}{h_j(t)} = e^{\beta(X_j - X_i)}$$

在风险不能满足比例性假定时，Cox 模型的估计可能会导致所有的参数，不仅仅是那些正在研究中的协变量的参数是有偏倚和无效的。比例风险假定应该用哈勒尔的 $\rho$ 来检查那些单个协变量，而用 Grambsch & Therneau 全局检验来检查非比例性（Box-Steffensmeier & Zorn, 2001; Therneau & Grambsch, 2000）。如果发现有非比例性的证据（而在大多数社会科学研究中，比例性与其说是一种常规，不如说是一种例外），那么潜在的成比例协变量应该与 ln（时间）交互或与其他适当的时间变换交互。诸如这样的交互允许每个被交互的协变量在冲突的风险上的效应单调地随研究的事件的持续期变化。放宽这一假定，将允许学者检验协变量的效应是否随时间变化，进而使我们得以更为精细地理解正在研究的现象。不仅如此，非比例性检验和它们所依据的残差，在普遍使用的用于分析持续时间数据的软件中越来越容易得到。

如果被估计的是参数模型，那么对选择的参数的分布的假定需要作检验，且比例风险假定可能仍然需要予以评估（例如，Weibull 模型也假定比例风险）。广义的 Gamma 分布是一种包容模型，包括若干种普遍使用的参数分布，因此也许有助于在竞争的嵌套模型之间作出裁定。因为参数模型是用最大似然估计的，且这些估计量的性质是众所周知的，一组标准的指数和统计量可直接用于参数建模框架，如使用似然比【Likelihood Ratio】检验和赤池信息化准则（Akaike information criterion, AIC）。然而，Cox 模型的最主要优点是不必对基线风险率的性质和形状作假定，因此对于社会科学家而言，在诸多建模策略中，它才是首选（Box-Steffensmeier & Jones, 1997）。

几乎所有的事件史模型的另一个潜在的假定都是所有的观察都会逐渐地经历感兴趣的事件。这一假定，可通过估计一个分离总体（split-population）模型（这些模型在生物统计学中也称为治愈模型，这一名称源于总体中一部分会被治愈之故）放松。以竞选献金时机研究为例，分离总体模型并不假定每一个政治行动委员会最终都会给予每一个政治候选人，而在刑事累犯研究中，模型并没有假定所有从前的囚犯最终都会再回到监狱。分离总体模型将那些将不会经历事件的观察的比例和那些描述经历事件

比例的风险率的特征的参数一起估计。这些模型允许协变量在事件是否发生及发生的时间上有不同的效应。例如,一个协变量可能在是否有献金上有正效应,而在什么时间献金上却有负效应。分离总体的对数似然的长处在于那些从未经历过事件的观察,只给函数的一部分提供信息。正因为如此,对数似然"分离"成了两个总体(Schmidt & Witte,1988)。

## 诊断

与传统的回归设置一样,事件史分析背景中的残差分析也是一种检查设定或模型的精确性的方法。为了检查模型的不同方面,事件史分析定义了各种伪残差,以将删截给残差定义增加的复杂性纳入考虑之列。伪残差除了在检验比例风险假定中使用之外,它们还能帮助研究者评估模型的拟合情况、协变量的函数形式和特定观察的影响。例如,鞅残差(martingale residuals)可被计算用来检验一个给定的协变量 $X$ 是否应该作为一个二次变量,还是许多其他可能的变量的一种线性进入。这些诊断方法在应用中应按规定使用,以确保模型的完整性。

## 单个和多个事件

除了研究事件一旦经历观察便离开了风险集合的单个事件发生之外,事件史分析也可以将多个事件纳入考虑之列。多个事件的事件史模型,考虑了跨事件的不独立性问题,因为忽略了相关问题可能会误导方差的估计,进而可能使系数的估计值出现偏倚。在这种情况下,我们考虑到"失败的"的几种类型中的一种,而诸如这样的过程则被称为竞争风险。例如,一位美国众议院的议员可能以各种各样令人感兴趣的方式离开众议院,我们应该对之有所认识,并把它们并入我们的模型。这些方式包括在初选中被击败、在大选中被击败、升职或退休等。我们可以期望风险率和协变量的效应将随离开的类型的不同而不同。

我们也可以考虑定序的多个(或重复的)事件。在重复事件过程中,被试者经历了同一类型的事件在一次以上,这样的情形在形形色色的公共卫生、劳动力和工业经济学、人口学和政治科学领域中并不罕见。没有对重复事件作解释意味着假定第一次、第二次、第三次和随后的事件在统计上是互相独立的。这是一个强的但通常是站不住脚的假定。条件事件时间间隔(或间隙时间)模型,将被应用于社会科学中的大多数重复事件场合。然而,我们之所以放弃其他的方法而选用这种方法,是因为该方法具有的性质,正是因为这种性质,重复事件才会发生(即相继或同时),且每一观察的相应的"风险集合"结构也应当会给我们提供最初选择这一模型,而非其他模型的动机(Box-Steffensmeier & Zorn,2002)。

## 例子和解释

在社会科学中,事件史分析应用的例子不胜枚举,且其涉及的领域还在不断扩张。这些例子包括失业期研究、和平、调查回应时间、刑事累犯、婚姻、公共政策项目实施和游说等。表 1 列出了一个典型的 Cox 比例风险估计,使用的是军事化冲突的数据(在 Cox 模型的估计中,对同分使用了 Efron 近似法。Breslow 近似法是首先被开发出来的近似法,但我们一般不推荐诸位使用它,然而在要相当长的时间才收敛时,精确似然选择给出的结果与 Efron 近似极为相似)。

表 1 冲突的 Cox 比例风险估计

| | $\beta$(S.E.) | $p$ 值 |
|---|---|---|
| 民主 | −0.439(0.100) | <0.001 |
| 增长 | −3.227(1.229) | 0.009 |
| 联盟 | −0.414(0.111) | <0.001 |
| 相邻 | 1.213(0.121) | <0.001 |
| 力量比率 | −0.214(0.051) | <0.001 |
| 贸易 | −13.162(10.327) | 0.202 |
| Wald 或 LR 检验 | 272.35($df$=6) | <0.001 |

注：$N$=20 448。

奥尼尔和拉西特广泛使用有关经济依附、民主和和平之间的关系的数据(Oneal & Russett,1997)被我们用来作为例证。该数据由 827 个对偶(即一对对国家,如美国和加拿大)从 1950—1985 年间的20 448个观察构成。我们设立了一个模型,把国际军事冲突风险率当作六个主要协变量(有些协变量会随时间变化)的函数:民主得分(两个国家的相对分数,值域从−1 到 1);经济发展水平(一个国家相对另一个国家经济发展速度的迟滞比,一个百分比量);对偶间联盟的存在(表示两国间是否存在联盟关系的虚拟变量);两国邻近情况(地理邻近情况的虚拟变量);两国的军事能力比(用以测量对偶间权力平衡的比率);对偶间双边贸易程度(用以测量相对较少贸易取向的国家双边贸易的重要性。它是双边贸易除以每个国家国内生产总值后的比值)。

自由主义理论认为,所有的变量除了相邻,都会减少争端的风险,而相邻则会增加争端的风险。表 1 底部的似然比(LR)检验显示,设定的模型比虚无模型更好(即虚无假设是设定的模型和没有独立变量的虚无模型之间在统计上并无显著不同)。所有的系数都在期望的方向,且除了贸易,所有的系数在统计上都是显著的。注意 Cox 模型是没有截距项的,或"上升"是跟随着协变量变化的(因此生存时间是在下降的),而负号则表明风险作为协变量的函数是在下降的。就这一模型而言,民主的负系数,−0.439表明并矢的民主降低了冲突的可能性,这就是说,并矢的民主导致更低的风险(和更长的生存时间)。博克斯-斯蒂芬斯迈耶和琼(Box-Steffensmeier & Jones,1997)用经历事件的风险变化的百分比来解释效应的影响(p.1434)。对于一个二分的自变量而言,经历事件的风险变化的百分比是

$$\frac{100 \times \left[ e^{(\beta_k \times 1)} - e^{(\beta_k \times 0)} \right]}{e^{(\beta_k \times 0)}}$$

负的系数产生值 $e^{(\beta_k \times 1)}$,它略小于 1,因此,产生了一个负的百分比变化。对于连续的自变量的解释则与之类似,即

$$\frac{100 \times \left[ e^{[\beta_k \times (x+\delta)]} - e^{(\beta_k \times x)} \right]}{e^{(\beta_k \times x)}}$$

这给出了自变量 $x$ 的一个 $\delta$ 单位变化的风险率百分比变化。这样,民主的一个单位的增长,将导致在任何给定的时间冲突的风险$[(e^{(-0.439)}-1)\times 100]=36\%$的相应下降。

事实上,军事化冲突的数据是由大量的重复事件描绘的,例如,英国和德国在第一次和第二次世界大战中相互打仗。博克斯·斯蒂芬斯迈耶和佐思(Box-Steffensmeier & Zorn,2001)用这些数据阐述重复事件持续期模型,并显示了由将重复冲突产生的相依性列入考虑之列所揭示的重要差别。

社会科学理论越来越关注变化的过程,与此同时,各种时态数据也变得越来越易于得到。事件史分析则是撬开这些研究元素的理想工具。近年来,这种技术的灵活性向多个事件的扩展,以及对有关感兴趣事件观

察的历史的整合,都使我们有理由期盼,事件史技术的使用和普及,必将大大推动社会科学的发展。

——Janet M. Box-Steffensmeier

(沈崇麟译校)

## 参考文献

Blossfeld, H.-P., & Rohwer, G.(2002). *Techniques of event history modeling: New approaches to causal analysis*. Mahwah, NJ: Lawrence Erlbaum Associates.

Box-Steffensmeier, J. M., & Jones, B. S.(1997). Time is of the essence: Event history models in political science. *American Journal of Political Science*, *41*, 336-383.

Box-Steffensmeier, J. M., & Zorn, C. J. W.(2001). Duration models and proportional hazards in political science. *American Journal of Political Science*, *45*, 951-967.

Box-Steffensmeier, J, M., & Zorn, C. J. W.(2002). Duration models for repeated events. *The Journal of Politics*, *64*(4), 1069-1094.

Cox, D. R.(1972). Regression models and life tables. *Journal of the Royal Statistical Society*, *Series B*, *34*, 187-220.

Hald, A.(1990). *A history of probability and statistics and their applications before 1750*. New York: John Wiley and Sons.

Kaplan, E. L., & Meier, P.(1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, *53*, 457-481.

Oneal, J. R., & Russett, B.(1997). The classical liberals were right: Democracy, interdependence, and conflict, 1950-1985. *International Studies Quarterly*, *41*, 267-294.

Schmidt, P., & Witte, A. D.(1988). *Predicting recidivism using survival models*. New York: Springer-Verlag.

Therneau, T. M., & Grambsch, P. M.(2000). *Modeling survival data: Extending the Cox model*. New York: Springer-Verlag.

# 事件抽样(Event Sampling)

事件抽样是指一类多样化的用于在自然、自发的背景下,研究个人的经验和社会过程的特定的经验方法。事件抽样法被设计用来得到一张思想、感情和行为在其日常生活中发生时的合理的明细表。这方面的例子包括经验抽样方法(experience sampling method, EMS),在使用这种方法时,被调查人会在一天某一随意时刻接到要求他们确切地描述自己此时此刻的活动的信号,这种方法还要求被调查人在每天结束时记日记,报告自己在那一天的经历,这种方法持续的时间一般在一个星期到一个月。在一个典型的 EMS 研究中,50 名大学生在一个星期中,每天被寻呼机通知 7 次。在接到信号通知时,那些对亲昵行为有着较高需求的学生(由投射个性试验评估确定的),更可能正在考虑的是他人和关系,更有可能正忙着和其他人进行交谈,而较少希望独处;如果他们正在交际中,则更可能报告正向的情感(McAdams & Constantian, 1983)。

事件抽样要求参与者监控并沿多维度,按预定的进度和由研究者定义的格式描述他们正在进行的活动。事件抽样有三个基本原理使之区别于其他通常使用的研究范式(例如实验室实验、抽样调查):(a)因为行为受到环境的影响,所以在自然环境中抽取行为这一点是很重要的;(b)整体性的回顾报告常常是有偏倚的,因为人们常常会受到回忆和概括很多过往事件的记忆力的限制;(c)重视看似平凡的日常经验,在能对

它们加以适当的审视时,便能给我们提供有价值的有关人类行为的洞见。一直在恰当地应用事件抽样的研究专题,包括情感研究、社会交往研究、痛苦研究、抽烟问题研究、压力和应对研究、运动研究、进食障碍研究、精神病力学研究、自我相关认知研究、个性研究、群际关系研究、药物治疗评估研究和治疗干预研究等。

大多数事件抽样都采用三种通用协议中的其中一种:依时回应(time-contingent responding),在使用这种方法时,参与者按固定的时间间隔(如每天、每小时)报告自己的经验;依事回应(event-contingent responding),这种方法要求每当一个事先定义的事件(如抽了一支烟和一个朋友交谈)发生时便要进行报告;依信号回应(signal-contingent responding),在使用这种方法时,只要某种设备(如寻呼机或一部编了程的袖珍计算机)给出了信号,参与者便要描述自己的经验。信号也许按一个固定或随机的时间表给出。尽管事件抽样源于简单的纸和笔的回应,但随着近年来电子记录设备(如个人数码助手和录音机)及移动的生理检测器的发展,不仅使它的有效性和可行性大增,且大大扩展了它的使用范围。

事件抽样不限于参与者的自我报告。例如,在学童的观察研究中,观察者可以给一个目标学童编码(如这个学童正在做什么事、当时在和谁互动、看得到的情感表达等)。编码的时间表可以是上述的任何一种(如分别按每隔 10 分钟、在冲突发生之后或依照一个随机的时间表)。

在一个典型的事件抽样中,一个研究者可能会得到两个星期的一系列有关青少年的瞬间情绪和行动的详细描述。然后,这些记录可以几种方式来使用,例如,依据该青少年在做什么及和什么人在做来确定其情绪有什么不同;分离个性因素或其他与情绪或活动模式关联的个人差异;或根据这些瞬间报告的概要来预测结果,例如,学业成绩、社交能力或健康状况等。不论具体研究的题目是什么,事件抽样研究者都专注于在自然环境中观察平常的行为,且都确信诸如这样的观察抓住了一些在行为和社会科学中的重大问题的重要线索。

——Harry T. Reis

(沈崇麟译校)

## 参考文献


Bolger, N., Davis, A., & Rafaeli, E. (in press). Diary methods: Capturing life as it is lived. *Annual Review of Psychology*.

Hektner, J. M., & Csikszentmihalyi, M. (2002). The experience sampling method: Measuring the context and the content of lives. In R. B. Bechtel & A. Churchman (Eds.), *Handbook of environmental psychology* (pp.233-243). New York: John Wiley & Sons.

McAdams, D. P., & Constantian, C. A. (1983). Intimacy and affiliation motives in daily living: An experience sampling analysis. *Journal of Personality and Social Psychology*, *45*, 851-861.

Reis, H. T., & Gable, S. L. (2000). Event-sampling and other methods for studying everyday experience. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (pp. 190-222). New York: Cambridge University Press.

Stone, A. A., Shiffman, S., & DeVries, M. (1999). Ecological momentary assessment. In D. Kahneman, E. Diener, & N. Schwarz (Eds.), *Well-being: The foundations of hedonic psychology* (pp. 27-38). New York: Russell Sage.

## 外生变量(Exogenous Variable)

外生变量是因果模型【Causal Modeling】或因果系统【Causal System】中的一个因素,它的值独立于系统中其他变量的状态,这就是说,它是一种值取决于研究之中的因果系统之外的因素或变量的因素。例如,降雨量对构建耕作过程和农作物产量的因果系统就是外生的。决定降雨量水平的是另外的因果因素,因此,对于天气模型而言,降雨量是内生的,但这些因素本身并非人们用来解释农作物产量水平的因果模型的一部分。

就内生变量而言,变量的状态有关一个特定模型的设定【Specification】和自变量【Independent Variables】之间的因果关系。至于外生变量,由定义可知,它的值完全在因果上独立于系统中的其他变量。因此,"外生"变量的类别与那些"纯粹内生的"和"部分内生的"变量形成了鲜明的对比。一个变量可以通过合并其他因素和因果关系而成为内生变量进入模型。

外生性有着因果和统计的解释。因果解释是基本的,并用对包含在模型中的其他变量的因素的因果独立性来定义外生性。统计学或计量经济学的【Econometric】概念则强调外生变量和模型中的误差项不相关。一般的回归【Regression】模型假定模型中的所有的自变量都是外生的。

——Daniel Little

(沈崇麟译校)

* 也可参见内生变量【Endogenous Variable】。

### 参考文献


Engle, R. F., Hendry, D. F., & Richard, J. F. (1983). Exogeneity. *Econometrica*, *51*, 277-304.

Pearl, J.(2000). *Causality*: *Models*, *reasoning*, *and inference*. Cambridge, UK: Cambridge University Press.

Woodward, J.(1995). Causation and explanation in econometrics. In D. Little(Ed.), *On the reliability of economic models*: *Essays in the philosophy of economics*(pp. 9-61). Boston: Kluwer Academic.


## 期望效应(Expectancy Effect)

参见实验期望效应【Experimenter Expectancy Effect】。

## 期望频数(Expected Frequency)

期望值等于某个统计值的平均数或均值,而期望频数则是一个事件发生次数的平均计数或平均数目。事件可能是一个单独的离散变量的一个类别、两个或更多个变量的交互分类的一格或某些其他的事件(例如,一个足球队在比赛中的进球数)。期望频数的计算方法是事件发生的概率乘以事件可能发生的次数的总数。例如,有一个从一个总体中抽取的容量为 $n=75$ 个人的随机样本。在该总体中,左撇子个体的概率为 $\pi=0.10$。在

样本中左撇子的人数可以为0~75的任何一个整数。然而,左撇子的期望或平均数则等于 $n\pi = 75 \times (0.10) = 7.5$。尽管频数总是整数,但期望频数一般却不是整数。

期望值是用于评估假设和模型拟合优度的检验统计量。两个常用的检验统计量是皮尔逊的卡方检验【Chi-Square Test】,$\chi^2 = \sum_{i=1}^{N} (f_i - F_i)^2 / F_i$ 和似然比统计量【Likelihood Ratio Statistic】,$L^2 = \sum_{i=1}^{N} f_i \log(f_i / F_i)$ ,式中的 $f_i$ 等于类别 $i$ 的观察数,$F_i$ 是类别 $i$ 的期望频数,而 $N$ 则是总类别数。如果虚无假设或模型为数据设定了一个概率(或若干个概率)值,例如,"左撇子的概率是0.10",那么期望频数就可以这样来计算,即 $F_i = n\pi_i$ ,式中的 $n$ 是总观察数。

事件的概率通常都是未知的,而是用数据估计的,估计的方法是计算样本比例或将模型与数据拟合。在将数据用于概率估计时,得到的期望频数称为估计的期望频数。例如,我们来看一下表 1 中的数据。该表是根据敦・科恩(Cohen,2001)提供的数据重新编排的。数据是英国利物浦 1902—1903 年的天花患者。这些患者分别用在婴儿期有否接种天花疫苗和患天花后是否死亡分类,然后将二者进行交互分类。假定疫苗对天花患者的病情没有影响,那么因天花而死亡的概率则可通过计算天花患者的死亡比例估计,也就是将总死亡数除以总个案数 $\hat{\pi}_{死亡} = 79/1\ 163 = 0.067\ 9$。由估计的死亡的期望频数可得到接种疫苗的死亡人数,它等于 $\hat{F}_{接种疫苗,死亡} = n_{接种疫苗}\,\hat{\pi}_{死亡} = 943 \times (0.067\ 9) = 64.055\ 9$。表 1 中,其他格子中的期望频数也可以计算。对于假设检验,则要用估计的期望频数 $\hat{F}_i$ 替代在皮尔逊卡方统计量或似然比统计量中的 $F_i$。

**表 1 英国利物浦的天花患者**
(1902—1903)

| | 因患天花死亡 | | 总计 |
|---|---|---|---|
| | 是 | 否 | |
| 婴儿期接种疫苗 | 19 | 924 | 943 |
| 未接种疫苗 | 60 | 160 | 220 |
| 总计 | 79 | 1 084 | 1 163 |

资料来源:Cohen,2001, p. 985。

——Carolyn J. Anderson
(沈崇麟译校)

## 参考文献

Agresti, A.(1996). *An introduction to categorical data analysis*. New York: Wiley.

Agresti, A.(2002). *Categorical data analysis* (2nd ed.). New York: Wiley.

Cohen, J.(2001). Smallpox vaccinations: How much protection remains? *Science*, *294*, p.985.

# 期望值(Expected Value)

期望值是一个随机变量【Random Variable】$X$ 的长期平均。更为正式地讲,期望值 $E(X)$ 是所有 $X$ 可能的值加权平均。对于离散【Discrete】变量而言,权是 $X$ 值的概率【Probability】$p(X)$,而平均数则用求和法得到:

$$E(X) = \sum_{x=-\infty}^{\infty} xp(x)$$

对于连续变量【Continuous Variable】而言,权是 $X$ 值的概率密度 $f(X)$,而平均数则用积分求得:

$$E(X) = \int_{x=-\infty}^{\infty} xf(x)$$

对于某些分布，积分或求和无法收敛。在这种情况中，期望则为未定义。尽管未定义的期望是相当少见的，但有一个众所周知的例子，发生在柯西分布中，该分布被定义为两个标准正态变量之比。

为了阐述期望值的计算，我们来看一下投掷一枚公平的硬币。如果硬币的面朝上，便令 $X$ 为 1，否则为 0。每个 $X$ 值都有着相同的概率 $p(X=1)=p(X=0)=0.5$，这样，使用离散变量公式的期望值为：

$$E(X) = 1 \times (0.5) + 0 \times (0.5) = 0.5$$

当 $X$ 是一个虚拟变量【Dummy Variable】时，期望值是 $X$ 为 1 的概率。在 $X$ 是定距【Interval】变量时，期望值是总体均值。

期望是一种线性运算：两个随机变量的加权和就是期望的加权和，即

$$E(aX + bY) = aE(X) + bE(Y)$$

式中的 $a$ 和 $b$ 就是权，而 $X$ 和 $Y$ 则是与两枚公平硬币关联的虚拟变量；如果硬币的面都朝上，那么每个变量的值就是 1。如果投掷两枚硬币，那么面朝上的期望数便是

$$E(X) + E(Y) = 0.5 + 0.5 = 1$$

正像和的期望就是期望的和一样，积的期望的积就是期望的积：

$$E(XY) = E(X)E(Y)$$

$X$ 的一般函数 $g$ 的期望就是 $g(X)$ 的加权平均。对于离散变量 $X$，权就是 $X$ 的值的概率 $p(X)$，而平均数则可通过求和得到。

$$E(g(X)) = \sum_{x=-\infty}^{\infty} g(x)p(x)$$

对于连续变量 $X$，权是 $X$ 值的密度 $f(X)$，平均数可通过积分求得：

$$E(g(X)) = \int_{x=-\infty}^{\infty} g(x)f(x)$$

这些公式很容易就可扩展到多变量的情形中，那时 $X$ 和 $Y$ 是随机变量向量，而 $g$ 则是变量 $X$ 的可能向量值的函数。

——Paul T. von Hippel

（沈崇麟译校）

## 参考文献

Rice, J. A. (1995). *Mathematical statistics and data analysis* (2nd ed.). Belmont, CA: Duxbury Press.

# 实验（Experiment）

在社会科学中，实验是社会科学家为了在一个或几个自变量和一个或几个因变量之间建立因果关系而采用的一种策略。自变量是一种由研究者操控的变量，它对研究者研究的因变量有因果影响。自变量有时也称为治疗变量或因素。因变量是受操控的自变量影响的观察的现象或量度。因变量有时也称为结果或效标变量。

在建立自变量和因变量之间的因果关系时，研究者应设计一种包括以下几个部分的实验：

- 操控自变量的量（在自变量为定量变量时）或水平（在自变量为定性变量时）。
- 用随机选择和对受试者随机分配治疗条件来控制无关（或混淆）变量。
- 仔细记录或观察自变量的变化。

第一个和第二个要求无论是在实验研究还是田野实验（参见田野实验【Field Experiment】）中都要得到实现；它们使实验研究策略区别于其他的研究策略，如准实验

研究、抽样调查或自然研究法。正因为如此，具有上述三个特点的实验有时也被称为真实验(Campbell & Stanley，1966)。

科多瓦和莱珀(Cordova & Lepper，1996)所做的有关小学生学习算术的运算次序法则的研究，便是一个很好的真实验的例子。所有的学习活动都发生在计算机化的环境中。首先，将同性别的学生随机地分配到控制条件组或四个实验条件组中的一个。在控制条件组中，学习材料抽象地进行讲解。在四个实验条件组中，同样的材料以有意义的和增强了视觉效果格式的方式作了情景化。四个实验条件中，一半学生的学习环境——作了个性化，另一半则仍然按照一般格式进行讲解。不仅如此。在每一个实验条组，都有一半的学生可在教学方面能对学习环境有所选择，另一半则不能。因此，这四个条件组是：(a)学习环境个性化，有选择；(b)学习环境个性化，没有选择；(c)学习环境非个性化，但有选择；(d)学习环境非个性化，且也无选择。总之，在研究中操控的自变量有三个：情景化、个性化和选择。因变量则有几个，包括学生内在的学习动机、学习期间投入的深度，在固定时期和时间获得的学习量，学生感知的自我能力和志向水平。三个自变量和多个因变量之间的因果关系通过三个步骤建立起来：(a)将受试者随机分配进控制或实验条件组；(b)控制各种变量，如受试者的学习成绩、性别和学习材料；(c)进行统计推论的统计理论。

的确，推论统计理论使社会和行为科学家能够理解用称为真实验方法收集的数据。在真实验中，因果关系的建立并非通过排除所有可能的备择解释或控制所有的混淆变量，而是借助概率。推论统计理论背后的逻辑是检验统计假设，检验的依据是数据。假设检验之后，便是以归纳的逻辑推理对潜在的总体进行推论(Maxwell & Delaney，1990)。引导我们进行实验研究的各种研究的原理和统计理论在相当大的程度上都是由卡尔·皮尔逊、罗纳德·费希尔和耶日·内曼提出和论证的。

社会科学家进行真实验或简称为实验的基本研究设计有三种(Kirk，1995)。它们是完全随机化设计、随机化分组设计【Randomized Blocks Design】和拉丁方【Latin Square】设计。基尔克把这些称为给所有的用于实验的研究设计构建模块(building blocks)。下面我们将对每一种设计逐一进行讨论。

## 完全随机化设计

这种研究设计包括一个自变量和一个因变量。自变量可以有两个或更多个水平(或治疗条件)，而因变量则总是连续的。如果把上面引用的科多瓦和莱珀(Cordova & Lepper，1996)的研究简化为只包括一个自变量，比如说只包括情景化，而观察到的因变量则是内在动机，那么简化后的研究设计便是一个真实的完全随机化实验，当然，被试者一开始便要被随机地分配给情景化或抽象条件。这样的简化研究布局看起来也许像表1。

表1

| 受试者 | 自变量 | |
|---|---|---|
| | 情景化条件 | 抽象条件 |
| $S_1$ | $Y_{11}$ | $Y_{12}$ |
| $S_2$ | $Y_{21}$ | $Y_{22}$ |
| ⋮ | ⋮ | ⋮ |
| $S_{15}$ | $Y_{1,15}$ | $Y_{2,15}$ |

以完全随机化设计收集的数据的统计线性模型是

$Y_{ij} = \mu + \alpha_j + \varepsilon_{ij}$($i = 1, \cdots, n$ 或本例中的15；$j = 1, \cdots, p$ 或本例中的2)

在上式中，$Y_{ij}$是第 $i$ 个受试者在自变量第 $j$ 个治疗条件的分数；$\mu$ 是所有受试者的总均值，因此是一个常数；$\alpha_j$ 是自变量的第 $j$ 个治疗条件的治疗效应；在代数上，它等于总体均值 $\mu_j$ 对总均值 $\mu$ 的偏差。对于所有在第 $j$ 个治疗条件的应试者的因变量分数是一个常数，受自变量的跨所有的治疗条件的所有的 $\alpha_j$ 之和为零的约束。$\varepsilon_{ij}$是和被随机地分配给自变量的第 $j$ 个水平的第 $i$ 个受试者关联的随机误差效应。它是一个随机变量 $\alpha_j$。

完全随机化设计的虚无假设（$H_0$）和备择假设（$H_1$）是

$$H_0: \alpha_1 = \alpha_2 = \cdots = \alpha_p = 0$$

$$H_1: \text{至少有一个 } \alpha_p \neq 0$$

为了检验虚无假设，数据被组织成条件间（between［$MS_b$］）的均方【Mean Squares】对条件内（between［$MS_w$］）的均方的比率。如果有关数据的正态性、方差齐性和独立性的假定得到了满足，那么这个比率服从 $F$ 分布。正态性假定说明在自变量的所有治疗条件中，数据的潜在分布是正态的。方差齐性的假定意思是所有的正态总体都有相同的方差。最后，独立性假定则是指随机误差 $\varepsilon_{ij}$彼此独立。无论条件内还是跨条件的都是如此。不能满足这些类似 $F$ 比率中的假定，会使第一类误差增大（Glass, Peckham, & Sanders, 1972）。上面给出的例子显然是一个最简单的完全随机化设计。它只有两个自变量的治疗条件，在这个例子中，检验统计量既可以是 $F$ 也可以是 $t$，因为要比较的是两个独立的组均值。在大多数社会科学的完全随机化实验的应用中，都会涉及三个或更多个治疗条件。

如果包括在研究中的自变量的治疗条件（例如情景化和抽象）是研究者在调查中仅有的感兴趣的治疗条件，那么研究者随后必定会对它们进行概括，这相应的自变量被看成一种固定因素，因而这样的设计则称为固定效应设计【Fixed-Effects Design】。适用于固定效应设计的统计模型是模型Ⅰ【Model Ⅰ】（Kirk, 1995）。然而，如果包含在研究中的治疗只是研究者希望研究的所有治疗的一个例子，那么相应的自变量则被视为是一个随机因素。该设计是一种随机效应设计，而模型Ⅱ【Model Ⅱ】（ANOVA）则是用于这样的数据的一种恰当的模型（Kirk, 1995）。

## 随机化分组设计

使用完全的随机化设计要求将应试者随机地分配给自变量的治疗条件。因此，治疗条件的应试者的样本（或子样本）被认为是独立于另外的治疗条件的应试者的。这种独立性特性将在进行统计检验时计算 $F$ 或 $t$ 统计量时考虑进去。对许多研究而言，例如那些涉及同一应试者的重复量度的研究而言，在不同治疗条件的应试者之间保持独立是不可能的。还有一些其他的研究（如家庭和婚姻研究）则宜将应试者进行配对（如夫妻或同班的同学），这样研究者就可以研究应试者之间的关系的动态。研究者也许会通过与某些相依量度有关的特性将应试者相配，故意引进相依性，以确保这些特性能得到控制。例如，在前述的简化的科多瓦和莱珀（Cordova & Lepper, 1966）研究中，我们可能希望根据与独立性量度（即内在动机）有关的从前的知识来将应试者相配。通过这种做法，那些归结于从前的知识的方差源便被计划安排进了研究之中，这样我们便可将它们与归结于治疗（即情景化）的方差隔离开来。让我们再进一步来假定，从前的知识差别有三个水平：高、中和低。一个控制从前的知识的恰当的研究设计便是随机化分组设计。表 2 阐明了这种设计的一些特点。

表 2

| 从前的知识 | 自变量 | | |
| --- | --- | --- | --- |
| | 应试者 | 情景化条件 | 抽象条件 |
| 高 | $S_1$ | $Y_{11}$ | $Y_{12}$ |
| | ⋮ | ⋮ | ⋮ |
| | $S_5$ | $Y_{15}$ | $Y_{25}$ |
| 中 | $S_6$ | $Y_{16}$ | $Y_{26}$ |
| | ⋮ | ⋮ | ⋮ |
| | $S_{10}$ | $Y_{1,10}$ | $Y_{2,10}$ |
| 低 | $S_{11}$ | $Y_{1,11}$ | $Y_{2,11}$ |
| | ⋮ | ⋮ | ⋮ |
| | $S_{15}$ | $Y_{1,15}$ | $Y_{2,15}$ |

从前的变量知识在技术上称为分组变量,它的水平(高、中和低)便是分组。在分组变量每一分组内的应试者(例如 $S_1,\cdots,S_5$)就从前的知识而言都是同质的。每一个分组的应试者被随机地分配给治疗条件(例如情景化条件或抽象条件)。每一个分组内的应试者被随机地分配给治疗条件和跨分组的有计划的差别,将确保应试者的特质差异在治疗条件中随机地分布,因而在分组变量上的差异在每一治疗条件中的表现将是等同的。因此,结果将不会因为在分组变量中的差异结果而系统偏倚。在这个意义上讲,随机化分组设计能将一部分分数方差隔离并归结于分组变量,而完全随机化设计则不能。

假定随机化分组设计的统计模型是:
$Y_{ij}=\mu+\alpha_j+\pi_i+\varepsilon_{ij}$($i=1,\cdots,n$ 或上例中的 3;$j=1,\cdots,p$ 或本例中的 2)
式中的 $Y_{ij}$ 是第 $i$ 个分组和第 $j$ 个治疗条件的分数;$\mu$ 是观察总体的总均值,因此是一个常数。

$\alpha_j$ 是第 $j$ 个治疗条件的治疗效应,在代数中,它等于总体均值($\mu_j$)离总均值($\mu$ 或 $\mu..$)的离差,对所有的在第 $j$ 个条件的所有观察的因变量分数则都是一个常数。像固定效应一样,受跨所有治疗条件的 $\alpha_j$ 之和等于零的约束;或像随机效应模型一样,治疗条件总体中的所有 $\alpha_j$ 之和等于零。

$\pi_i$ 是第 $i$ 个总体的分组效应且等于第 $i$ 个总体的均值($\mu_i$)离总均值($\mu..$)的离差;对于所有在第 $i$ 个分组的观察的因变量分数,它都是一个常数,在潜在的总体中正态分布。

$\varepsilon_{ij}$ 与第 $i$ 个分组和第 $j$ 个治疗条件关联的观察的误差效应;在第 $i$ 个潜在总体和第 $j$ 个治疗条件的潜在总体中正态分布,且独立于 $\pi_i$。

随机化分组设计的虚无假设($H_0$)和备择假设($H_1$)等同于完全随机化设计。为了检验虚无假设,数据被组织成均方治疗条件($MS_t$)与均方残差的比率($MS_r$)的形式。均方残差是均方残差和与它的自由度的比率,其中均方残差和是内平方和($SS_{内}$)与可归结于分组变量的平方和之差。如果关于数据的正态性、等方差性和独立性假定都得到了满足,那么该比率服从 $F$ 分布。

上面给出的例子是一种只有两个治疗条件的最简单的随机化分组设计。在这种情形中,检验统计量既可以是 $F$,也可以是 $t$,如同对两个相依(或相配)的组均值进行比较一样。在社会科学中,大多数随机化分组设计的应用都涉及三个或更多个的治疗条件。

如果包含在研究中的自变量治疗条件(如情景化和抽象)都被看成是固定的,且分组变量也同样被看成是固定的,那么这样的设计就是固定效应设计。用于固定效应设计的统计模型是模型Ⅰ(Kirk, 1995)。然而,如果分组变量被认为是一个随机因素【Random Factor】,那么与之对应的设计就是一种混合设计【Mixed Design】,这个模型就是模型Ⅲ【Model Ⅲ】(ANOVA)。如果包括在研究中的模型治疗条件是代表所有治疗条件的一个样本且相应的自变量因此是一个随机因素,那么这种设计就是一种随机效应设计,这样模型Ⅱ便是适用于这样的数据的恰当的统计模

型(Kirk, 1995)。

在有一个以上的应试者被分配给分组和治疗条件的组合时,这种设计就称为广义随机化设计。这里给出的例子便属于这种情况。或者,同一个应试者(或观察)可以作为一个分组,因而受一个随机序列中的所有治疗条件的支配。有些作者喜欢将分组设计的这种特殊的变种称为重测量度设计【Repeated-Measures Design】。有关分组是怎样形成的,以及与这种设计关联的特有的统计假定,以及与它对应的统计检验可参见词条目"分组实验设计(Block Design)和随机化分组设计(Randomized-Blocks Design)"。

## 拉丁方设计

拉丁方设计是一种实验设计或准实验设计【Quasi-Experiment】,它可以隔离并将一部分方差分数归结于两个分组变量,而随机化分组设计只能控制一个分组变量。

与魔方非常相像,所有的拉丁方设计都是正方形的(2×2,3×3,4×4 等),这就是说,两个分组变量和自变量都有相同的水平(或条件)数。我们假定在上述简化的科多瓦和莱珀研究中,想要控制的第二个变量是性别,它有两个水平。自变量也有两个水平:情景化和抽象。因此,第一个分组变量的从前的知识必须重新定义成只由两个水平组成:高和低。表 3 阐明了这种 2×2 的拉丁方设计。

**表** 3

| 从前的知识 | 性别 | |
|---|---|---|
| | 女性 | 男性 |
| 高 | 情景化条件 | 抽象条件 |
| 低 | 抽象条件 | 情景化条件 |

注意,在表 3 中治疗列已在正方形中,而两个分组变量则分别为行和列变量。治疗条件确实在每一行和每一列都只发生了一次。

上面的拉丁设计的统计模型是:

$Y_{ijkl} = \mu + \alpha_j + \beta_k + \gamma_l + \varepsilon_{合并}$($i = 1, \cdots$ 上例中的 $n$; $j$, $k$ 和 $l = 1, \cdots, p$ 或本例中的 2)

式中的 $Y_{ijkl}$ 是在第 $j$ 个治疗条件、第一个分组变量(例如从前的知识)的第 $k$ 个水平和第二个分组变量(例如性别)的第 1 个水平中的应试者分数。

因此,$\mu$ 观察总体的总均值,是一个常数。

$\alpha_j$ 是第 $j$ 个治疗条件的治疗效应,用代数方法计算,它等于总体均值($\mu_{j\cdot\cdot}$)离总均值($\mu$ 或 $\mu_{\cdots}$)的离差;对于所有应试者因变量分数它都是一个常数,受跨所有的治疗条件的所有 $\alpha_j$ 之和为零的约束,如同在固定效应设计中一样;或受治疗总体中的所有 $\alpha_j$ 之和为零的约束,如随机效应设计。

$\beta_k$ 是第 $k$ 个总体的分组效应,它等于第 $k$ 个总体的均值($\mu_{\cdot k}$)减总均值($\mu_{\cdots}$)的差。对于在第 $k$ 个分组的应试者因变量分数它都是一个常数。在它的潜在总体中服从正态分布。

$\gamma_l$ 是第 $l$ 个总体的分组效应,它等于第 $l$ 个总体均值($\mu_{\cdot\cdot l}$)同总均值($\mu_{\cdots}$)的差;对于在第 $l$ 个分组中的应试者因变量分数,它都是一个常数,在它的潜在总体中服从正态分布。

$\varepsilon_{合并}$是与 $Y_{ijkl}$ 关联的合并误差效应,用代数方法计算,这个概念等于观察分数离所有的模型参数(如 $\mu$,$\alpha_j$,$\beta_k$ 和 $\gamma_l$)的离差,且根据它的定义,其在统计上独立于所有的模型参数。

拉丁方设计的虚无假设($H_0$)和备择假设($H_1$)也和完全随机化设计相同。为了检验虚无假设,数据被组织成均方治疗条件($MS_t$)与均方合并误差($MS_{合并}$)的比率,均方合并误差是一个比率,是均方合并误差之和与它的自由度之比。而均方合并误差之

比则是内平方和($SS_{内}$)减去归结于两个分组变量的平方之差。如果有关数据满足了正态性、方差齐性和独立性假定,那么这一比率服从 $F$ 分布。此外,$F$ 检验还假定两个分组变量和自变量(也包括与治疗条件)之间不存在交互。为了检验这个交互假定,一个拉丁方设计的每一格中必须有两个或更多的观察(即 $n \geq 2$)。否则,这个无交互假定只是对数据的假定。

如果包括在一个研究中的自变量治疗条件(如情景化和抽象)被看成是固定的,且两个分组变量也同样被看成是固定的,那么该设计就是一种固定效应设计。用于固定效应设计的统计模型是模型Ⅰ(Kirk, 1995)。然而,如果分组变量中有一个被看成是一种随机因素,那么相应的设计就是一种混合效应设计,那么适用的模型就是模型Ⅲ。如果包括在研究中的治疗条件只是所有治疗条件的一个样本,那么对应的自变量便因此而是一个随机因素,设计也就因此而成为一种随机效应设计,所以适用这样的数据的统计模型便是模型Ⅱ。有关拉丁方设计的更多的阅读资料,可参阅基尔克(Kirk,1995)的《拉丁方设计》(*Latin Square Design*)或马克斯韦尔和德莱尼(Maxwell & Delaney,1990)的著作。

上面介绍的三种设计和对应的统计模型,都假定数据的收集和测量是没有误差的。如果在数据中存在测量误差,那么在计算 $F$ 统计量完成之后,就必须考虑信度问题(Cleary & Linn, 1969; Levin & Subkoviak, 1977)。不仅如此,任何在试验中,在自变量和一个或多个因变量之间建立起来的因果关系的效度都有可能受到一些因素的威胁。有关对外部、内部和统计结论的效度有威胁的问题的详细介绍,请参阅坎贝尔和斯坦利(Campbell & Stanley,1966)及基尔克(Kirk,1995)的著作。

在概率理论领域,实验【Expriment】是一个导致其发生概率是确定的单一结果的过程或程序。一个实验可以是实验室研究(如参与的学生在协作学习中提高他们的社会技巧)或一个取自某一总体的观察样本的行为的某一方面的过程(记录从一个 50 人的晚宴中随机抽取的两名客人的出生月份)。实验和结果二者(社交能力的提高或两名客人出生在同一个月份)都应当被很好地定义,以使研究者能制订确定在所有的可能结果中得到特定结果的概率规则。确定概率的路数既可以基于主观-人格的框架(subjective-personalistic framework),也可基于逻辑或经典的框架,还可以基于经验比较或频数的框架。

——Chao-Ying Joanne Peng

(沈崇麟译校)

＊也可参见分组设计【Block Design】、田野实验法【Field Experimentation】、拉丁方【Latin Square】、准实验【Quasi-Experiment】、信度【Reliability】、效度【Validity】。

## 参考文献

Campbell, D.T., & Stanley, J. C.(1966). *Experimental and quasi-experimental designs for research*. Chicago: Rand McNally College Publishing.

Cleary, T. A., & Linn, R. L. (1969). Error of measurement and the power of a statistical test. *The British Journal of Mathematical and Statistical Psychology*, *22*(1), 49-55.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Chicago: Rand McNally College Publishing.

Cordova, D. I., & Lepper, M. R. (1996). Intrinsic motivation and the process of learning: Beneficial effects of contextualization, personalization, and choice. *Journal of Educational Psychology*, *88*(4), 715-730.

Fisher, R. A.(1971). *Design of experiments*. New York: Hafner Press.(Original work published 1935.)

Glass, G. V., Peckham, P. D., & Sanders, J. R. (1972). Consequences of failure to meet assumptions underlying the analysis of variance and covariance. *Review of Educational Research*, *42*, 237-288.

Kirk, R. E.(1995). *Experimental design: Procedures for the behavioral sciences*(3rd ed.). Belmont, CA: Brooks/Cole.

Levin, J. R.,& Subkoviak, M. J.(1977). Planning an experiment in the company of measurement error. *Applied Psychological Measurement*, *1* (3), 331-338.

Maxwell, S. E.,& Delaney, H. D.(1990). *Designing experiments and analyzing data: A model comparison perspective*. Belmont, CA: Wadsworth.

Vogt, W. P. (1999). *Dictionary of statistics and methodology: A nontechnical guide for the social sciences*(2nd ed.). Thousand Oaks, CA: Sage.

# 实验设计(Experimental Design)

## 实验的典型特征

实验的两个典型特征使它与其他的社会探究方法相区别。实验的第一个典型特征是都要设计一种有计划的干预。与研究者试图查探在推断的因果中自然发生的波动效应的被动观察不同,实验则在环境中创建了一种干扰,进而去探查它的结果。例如,一个福利改革的观察研究,可能去观察失业率是否随着时间或项目要求的辖区的变化而变化。相反,实验调查研究可能将项目的参与者分配给不同的需求集合,以考察它们是否对失业有影响。

实验的第二个典型特征是随机分配【Random Assignment】。尽管实验【Experiment】一词在日常话语中的使用是比较宽泛的,泛指任何形式的干预,但在社会科学中,它却专指研究观察单位被随机地分配给治疗和控制条件的研究。因为社会科学家无法创建相当于物理实验一样的去除了外部因果关系的实验室,他们只能改而依靠随机分配。正如弗希尔(Fisher,1935)在其经典著作《实验设计》(*The Design of Experiments*)中所指出的一样,这种方法是唯一能保证治疗组和控制组具有可比性的方法。只要充分实现了随机化,就可以肯定治疗和控制组的差别,仅仅是因为机会的结果。

因此随机化使研究者能够对来自与研究的干预对照的组间偶然差异的,治疗和控制组之间的任何干预后差异作出精确的统计陈述。在非实验的社会科学研究中,一般难以获得精确的概率陈述,因为那时自变量都是由一个未知的过程生成的。除了起因于有限样本量的通常的统计不确定性之外,非实验研究还涉及大量不确定性,例如,观察到的两个变量 $X$ 和 $Y$ 之间的相关,是否反映了 $X$ 对 $Y$ 或 $Y$ 对 $X$ 的因果影响;或者反映了第三个变量 $Z$ 对 $X$ 和 $Y$ 两个变量的因果影响?

## 实验设计面临的问题

尽管从原则上讲,实验给我们提供了进行统计推论的最可靠的工具,但是在实践中实验会遇到一些问题。在一个实验未能将特定的诱因体的真正效应隔离开来时,就会出现内在效度【Internal Validity】问题。例如,拉纳克郡牛奶实验(Lanarkshire milk experiment),这是一个早年设计用来检验配给学校的牛奶是否提高了学生的成长率的实验。该实验未能取得成功,因为教师将在控制组中的体重不足的学生重新分配,以使他们能接受牛奶供应。这种治疗和控制组的对比同时反映了牛奶和教师重新分配的

异常行为二者的效应。

在实验进行的环境或参与实验的人在一些重要方面不同于研究者试图概括的地方和总体时，便会产生外在效度【External Validity】问题。在实验室实验【Laboratory Experiment】试图用大学生作为应试者来模拟经济或政治环境时，便常常会出现这种问题。在试图解决外在效度问题时，社会科学家越来越多地采用实地实验【Field Experiment】或在真实世界中进行实验。这样的研究已经考察了选民动员活动、学前教育项目、鼓励按法纳税的方法（methods for encouraging compliance with tax rules）和许多其他干预的实效性。然而种种实际的约束限制了实地实验的实施范围。社会科学家缺乏资源和操控，如立法机构或总体的宗教信仰这样的大规模致病因素的权威。

所谓外在效度问题是一个同质治疗效应问题。即所有的受试者是否都受到了给定的干预的同等影响？如果答案是“是的”，那么研究机遇的范围便会扩大，因为我们可以研究这样一些情景，在这样的情景中只有那些人中的某些被分配到了治疗组的人才实际得到了治疗。例如，如果在一个选民动员运动中，对它所接触的每一个人都有相同的影响，但是它却只找到了它想要寻找接触的人的一半，那么它的影响强度将是纯粹的治疗组和控制组的比较的两倍。换句话说，同质治疗假定使我们得以轻易地从那些实际接触了治疗的人，外推到研究者寻求治疗的那些人。

治疗效应是否实际上同质是一个经验的问题，与大学生是否与那些大学外面的人一样易受社会影响这个问题不同。同质性和外在效度的关注点在于强调在不同的场景和总体中复制实验发现的重要性【Replicating】。

伦理限制是一种更进一步的约束。实验干预可能对参与的应试者和更为广泛的社会有不利的影响。此外，在发明可行的干预的实际约束之外，研究者更多的是依照某些程序性保障保证应试者既不要受到伤害，也不要受到强迫。出于这些考虑，我们一般不再考虑进行深远的涉及社会和经济政策的实验，但诚如唐纳德·坎贝尔（Campbell，1969）所指出的一样，不进行这样的实验的社会代价也必须在我们考虑之列。政府、公司和组织可以继续干预这个世界，但问题在于他们的干预是否以能产生有用的知识这样一种方式来规整。

## 实验和科学进步

尽管有这些限制，但实验仍不失为一种进行因果判断的好方法。一个单独的、设计精当的实验，若在真实世界的情景中进行，且数量足够多，那么它所产生的统计上精确的结论，将远胜于大量以观察数据为根据的研究。

实验的某些诱惑力在于它的简洁透明。与非实验数据分析相比，实验结果的分析，常常略为超出基本统计方法的范围，且统计方法的选择也几乎与结果承载的材料无涉。实验设计在很大程度上已经规定了将要采取的数据统计分析的方式；按事先计划的分析方式进行分析，而不在事后进行分析方法的选择，以免结果有朝向某一特定方向的偏倚。实验程序不仅能导向更为清晰的因果推论，而且也可以使分析者免于数据挖掘的道德风险。

——Donald P. Green

（沈崇麟译校）

* 也可参见实验【Experiment】。

## 参考文献

Campbell, D. T. (1969). Reforms as experiments. *American Psychologist*, *24*, 409-429.

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Boston: Houghton Mifflin.

Fisher, R. A. (1935). *The design of experiments*. New York: Hafner Publishing.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Wilmington, MA: Houghton Mifflin.

## 实验者期望效应(Experimenter Expectancy Effect)

实验者期望效应是在科学探索中的人为的效应或误差源之一(参见调查效应【Investigator Effect】),它特别指在实验者的假设或研究结果期望上的非预期的效应。

在科学中,某些对于研究结果的期望几乎是永恒不变的。社会科学家通常都与其他科学家一样,在进行研究时都特别想考察一些有关事物本质的假设或期望。在社会和行为科学中,调查者持有的假设【Hypothesis】可能会导致他们在不经意间改变研究参与者回答问题的方式,使他们更有可能给出有利于肯定调查者假设或期望的回答。于是我们说调查者的假设好比是一个自我应验的预言。有人预言一个事件,而有关这一事件的预言然后就改变了预言者的行为方式,以使预言更有可能实现。科学的历史用聪明的汉斯马(horse Clever Hans)这一事例作为一个主要的例子记载了这样一种现象的发生(Pfungst, 1911/1965)。

第一批实验专门设计用于调查实验研究雇来的研究参与者结果的实验期望效应。心理学领域的研究生和高年级的大学生被招聘来从普通心理学学生中收集数据。实验者对研究参与者出示一系列面部的照片,并要求参与者给照片上反映出来的成功或失败打分。让一半实验者,期望他们随机选择的研究参与者在给照片打分时,有更多的成功的人;而其余一半的实验者则给出了相反的期望——期望他们的研究参与者给出的分数有较少的成功者。尽管所有的实验者都被要求做完全是标准化的实验,且向所有的参与者都宣读了同样的打印的说明,但那些曾期望给面孔的打分有较多的成功人士的实验者,用他们的随机分配的参与者得到了这样的打分。而那些曾被导向期望结果相反方向的实验者则易于得到反向的结果。

这些结果用其他的人类研究参与者重复了许多次。他们也用动物研究应试者重复了实验。在这些实验的第一个实验中,雇用的实验者被告知,他们的实验室正和另一个开发出了在迷津学习中比较聪明的老鼠和在迷津学习中比较愚笨的老鼠的遗传品系的实验室合作。实验者的任务只是简单地记录聪明的老鼠和愚笨的老鼠在迷津学习中的成绩。一半的实验者被告知,给他们分配了聪明的老鼠,而其余的实验者则被告知,分配给他们的都是愚笨的老鼠。实际上没有一只老鼠真的被培育成迷津学习的聪明的老鼠,也没有一只老鼠被培育成迷津学习的愚笨的老鼠。而实验者则是完全随机地被告知,分给他们的是什么样的老鼠。尽管事实上,据说聪明的和愚笨的老鼠之间的差别只存在于实验者的脑子中,那些相信自己的老鼠是聪明老鼠的实验者,从他们的老鼠那里得到了较好的成绩,而那些认为自己的老鼠是愚笨的实验者,则从他们的老鼠那里得到了较差的成绩。在用斯金纳箱(Skinner boxes)而非迷津重复这一实验时也得到了基本上相同的结果。

——Robert Rosenthal

(沈崇麟译校)

*也可参见皮格马利翁效应【Pygmalion Effect】。

## 参考文献


Pfungst, O. (1965). *Clever Hans*. Translated by C. L. Rahn. New York: Holt, Rinehart and Winston. (Original work published 1911.)

Rosenthal, R. (1976). *Experimenter effects in behavioral research: Enlarged edition*. New York: Irvington Publishers, Halsted Press Division of Wiley.

Rosenthal, R., & Jacobson, L. (1968). *Pygmalion in the classroom*. New York: Holt, Rinehart and Winston.

Rosnow, R. L., & Rosenthal, R. (1997). *People studying people: Artifacts and ethics in behavioral research*. New York: W. H. Freeman.


# 专家系统(Expert Systems)

专家系统(ES)也称为基于知识的系统,是用于有限知识域的具有复杂的决策和推理能力的人工智能【Artificial Intelligence, AI】计算机程序。专家系统一般来说都使用常情诠释【Abduction】(指向推论的推理),它们不只是智慧型的数据库,因为它们可以模拟人类的思想。对社会科学家来说,专家系统提供了给不同的理论框架建模、检验理论的假设、进行模拟【Simulation】的办法。专家系统也在教学、研究设计、问题求解、鉴别、诊断和控制中有广泛的运用。

通常 ES 由四个部分组成:(a)知识库本身,包括某个领域的陈述性知识("材料")和程序性知识(如何使用"材料")的主体;(b)推理引擎,提供推理的工具;(c)知识获取界面,规定了获取知识的模式和规则形式;(d)用户界面,提供一个直观的,像人类一样的终端用户接口。通过专家系统给问题建模,与通过电话咨询由有真人专家来直接解决这一问题是等价的。

举例来说,表 1 中所展示的是基于规律的 ES 对怀孕妇女进行第一次产前检查的时间早还是晚所做的预测的简单解释。这是从对做过第一次产前检查的妇女的访谈田野笔记【Fieldnotes】中建构出来的。它的操作可以与多元回归分析【Multiple Regression Analysis】相比较,但更容易解读。

表 1

| | |
|---|---|
| 规则 1 | **如果**受到有影响力的他人的鼓励=是,<br>**那么**第一次产前检查就早 |
| 规则 2 | **如果**受到有影响力的他人的鼓励=否,<br>**或**受到母亲的鼓励=是,**那么**第一次产前检查就晚 |
| 规则 3 | **如果**受到有影响力的他人的鼓励=不确定(根据访谈资料)<br>**且**早期的症状=是,**那么**第一次产前检查就早。<br>**否则**(即,如果否),第一次产前检查就晚 |

尽管 ES 经常都用人工智能(AI)语言,诸如 C++,Prolog 或 Lisp 来写,但许多 ES 还是用一种称为壳体(shells)的程序设计环境来开发的。一个典型的壳体提供了专家系统四个关键部分中的三个部分——推理引擎、知识获取界面和用户界面——只有知识数据库是由开发者提供的。

社会科学家发现,在系统阐述和检验自己的认识时,建立一个 ES 是一种有用和发人深省的经验,因为 ES 需要专属的假设和试探法(专家用来操控知识的非正式的经验法

则)。ES 允许基于溯因的、基于归纳法【Induction】和概率性【Probabilistic】的推理,因此它并非只是一个基于规则的模型【Model】。

相关的方法还包括数据挖掘、归纳法,以及模糊系统。在这里,类包含(class-inclusion)可能会有模糊的边界。其缺点就是不能处理没有预料到的关系以及存在系统维护的问题,尽管专家系统的壳体比 C++ 等语言更加自文档化(self-documenting)。

本弗、布伦特和弗比(Benfer, Brent, & Furbee,1991)给社会科学家提供了一个易于理解的入门材料。在线资源很丰富,含可索取的一般信息和入门指南(有一般的商业信息,以及包括免费软件和共享软件在内的资源)。

——N. Louanna Furbee

(杨可译　沈崇麟校)

## 参考文献

Benfer, R. A., Brent, E. E., Jr., & Furbee, L. (1991). *Expert systems*. Newbury Park, CA: Sage.

# 已解释方差(Explained Variance)

参见 *R* 方【*R*-Squared】。

# 解释(Explanation)

一个事件或一种规律性的科学解释是一种论证,其目的在于显示在世界的其他性质给定时,事件或规律性为什么会发生。在既定情况和相关的所依法则或规律性存在时,为什么这个结果是必然或可能的。科学解释的最为一般的模式是"覆盖律解释"(covering-law explanation)。一个覆盖律解释是一个包含一个或几个一般规律、一个或几个特别的陈述(边界条件)和被解释现象的陈述的演绎推导论证。这样的解释被设计用来显示在初始条件和性质规律给定时,一个结果为什么是必然的。一个覆盖律解释把现象归入一般的规律。概率性解释有着与之相似的逻辑。概率性解释确认一个或几个统计规律,并把现象归入这些规律:给定概率规律 $L_i$,$O$ 的概率是 $P$。这些方法的科学解释的逻辑把科学规律或要解释的性质的作用放在了首要位置。使用这一方法,在相关性质的规律给定时,在我们已显示一个结果是如何必然时便已对它作出了解释。

科学解释的另一种方法发端于世界是一个因果过程和因果机制系统这样一种观点。在使用这种方法时,我们对因果机制和导致结果发生的力量作出解释时便对一种现象也作出了解释。因果解释始于因果机制的确立,然后再通过它,由初始条件引出待解释项。这两种方法是有关联的,因为因果机制的存在也意味着,可在覆盖律内起作用的类似规律概括的有效性。在我们问一个结果的解释时,我们想知道事件为什么会发生?托勒密关于天空中行星的观察位置的解释,是以在一组类似规律的概括(循环和本轮)包容了行星运动为依据的。这种解释是不能令人满意的,因为它依一种导致一些天文观察的因果过程的伪概念为根据(以地球为中心,而非以太阳为中心的行星运动)。

——Daniel Little

(沈崇麟译校)

## 参考文献

Achinstein, P. (1983). *The nature of explanation*. New York: Oxford University Press.

Hempel, C.(1965). *Aspects of scientific explanation, and other essays in the philosophy of science*. New York: Free Press.

Kitcher, P., & Salmon, W. C. (Eds.). (1989). *Scientific explanation* (Minnesota Studies in the Philosophy of Science, Vol. 13). Minneapolis: University of Minnesota Press.

Little, D.(1991). *Varieties of social explanation: An introduction to the philosophy of social science*. Boulder, CO: Westview.

Miller, R.W.(1987). *Fact and method: Explanation, confirmation and reality in the natural and the social sciences*. Princeton, NJ: Princeton University Press.

Salmon, W. C.(1984). *Scientific explanation and the causal structure of the world*. Princeton, NJ: Princeton University Press.

# 解释变量(Explanatory Variable)

解释变量是一种分析者用来解释一种现象的因素。在回归【Regression】模型中，它也称为自变量【Independent Variable】、回归量(regressor)、输入变量、右侧变量或$X$变量。

——Tim Futing Liao

(沈崇麟译校)

# 探索性数据分析(Exploratory Data Analysis)

探索性数据分析(有时缩写为EDA)是一种数据分析的方法。它令数据本身自行揭示自己的潜在结构，从而使研究者对数据有所“感觉”。要达到这一目的，在很大程度上有赖于图标和展示，因为目测给我们提供了一种对数据的特别洞察，但与此同时，它也使用了许多数字技术。目测和数字技术二者的重点集中在寻找使我们得以更有效地检验理论和假设的知识。假定探索性数据分析代表理解数据的一般哲学，那么这些技术只是一种工具，而非目的本身。

探索性数据分析的哲学有两个重要的特性：开放性和怀疑论(Hartwig & Dearing, 1979, pp. 9-12)。分析应当是对未曾预料到的、超出计划的模型和期望的那些数据模式持开放的态度；而对那些可能会隐匿或曲解数据中最有信息量的方面的概括数字则应持有怀疑的态度。

## 历史发展

数据分析的开创性工作始于约翰·图基(Tukey,1977)，弗雷德里克·莫斯特勒和约翰·图基(Mosteller & Tukey, 1977)。图基(Tukey,1977, p. vii)注意到，用以确认样本【Sample】统计量在什么样的精确程度上可用来推论总体【Population】参数的验证性数据分析一直主导着20世纪的统计研究。他认为，探索性和验证性数据分析强调的是不同的原理，但是二者都应当为实际的研究者所用。探索性数据分析的目的在于对居主导地位的统计方法作补充，他认为研究者不可以没有验证性数据分析，但是数据分析却不必从它开始。他的主要观点是(Tukey, 1977)：“重要的问题在于在学会自己做得有多好之

前,先明白自己究竟能做什么。”(p. v)

我们来看一下验证性数据分析和探索性数据分析的不同之处(NIST/SEMATECH, 2003)在于:

- 验证性方法始于一个预设的模型,并用来自数据分析得到的参数来评估模型,而探索性方法则始于数据分析,以推测模型和设定它的假定【Assumption】。
- 验证性方法是正式和严格的,旨在评估模型,而探索性方法则不那么正式,且比较灵活,旨在发现数据含有的意义。
- 验证方法在数据上建立模型,而探索性方法则让数据提出一些适当的模型。
- 验证性方法筛选数据,将信息减少到少数参数,而探索性方法则依靠反映数据所有部分的那些方法。
- 验证性方法要求有强的假定,而探索性方法只做少数几个假定,或试图在模型检验前确认假定。
- 验证性方法更有鉴定力,以精确地检验假设【Hypothesis】,而探索性方法则有更大的普遍性,但灵敏性较低。

这两种方法代表理想的类型,在一个优良的研究中,两种方法是并重的。图基和其他一些学者一贯反对那种认为,只是对数据作描述和探索的做法,其层次低于正式的假设检验和模型验证的观点。今天,虽然大多数经验的社会科学研究都依靠验证性的模型,例如,使用回归【Regression】和方差分析来检验理论,但是探索性技术也已经成为一个通向验证性模型研究过程的一个基本部分。

## 实例

图基(Tukey, 1977)及莫斯特勒和图基(Mosteller & Tukey, 1977)的著作,给我们提供了许多用于探索性数据分析的方法和统计量,甚至还有很多新的术语。少数几个有关这些方法和统计量的例子,以及它们对实现探索性数据分析的贡献,可给研究者阐明它们的潜在用途。

为了揭示隐藏的数据结构,探索性数据分析使用图表和图形来帮助研究者注意到他们从未想看到的东西。一张茎叶图【Stem-and-Leaf Display】和一张盒须图(参见盒形图【Boxplot】)每一张都以直观的方式描述一个单变量【Variable】的值的完整分布【Distribution】,而散点图【Scatterplot】则描述了含有所有的变量值的两个变量之间的关系。视觉再现能在很大程度上确认任何对正态性的偏离【Deviation】(在单变量中)和对线性的偏离(在两个变量的关系中),从而帮助我们对通常用于最小二乘【Least Squares】回归的假定进行评价。

为了解决非正态性和非线性【Nonlinearity】问题,探索性数据分析寻找一些重新表述变量量度量表【Scale】的方法。重新表达在保持顺序或次序不变的同时,改变变量值之间的距离。取变量的幂、根和对数,使分布正态化及关系直线化,使变换后的变量更适于做多变量分析。使顺序变量变得平滑的技术同样也能帮助我们揭示数据中的潜在模式。

为了最好地对分布和关系进行概括,探索性数据分析都有赖于稳健【Robust】计算和抗性量度。反映数据中的大多数使抗性量度对偏差个案分析【Deviant Case Analyses】不那么敏感。这样的量度经常以中位数【Median】而非均值;以绝对值而非平方值,以及以最小绝对离差而非最小平方离差为依据。各种稳健的回归技术进一步提供了使模型变得不太敏感和降低离群个案的不成比例影响的方法。

为了评估模型的恰当性,探索性数据分析对离群值【Outlier】和残差【Residual】给予了更多的关注。变量之间的关系的最小平方分析都辅以用散点图和点对回归线的偏离的视觉检查。许多数字技术虽然也能确认多变量背景中的大异常值和残差,但是它

们并不能完全取代从带有数据的模型的视觉对比中得到的好处。其次，关注异常值和残差又有助于我们确定如何重新表述变量、重新设定模型，以更好地描述关系，进而得到对正在调查的专题新的洞见。

——Fred C. Pampel

（沈崇麟译校）

## 参考文献


Hartwig, F., & Dearing, B. E. (1979). *Exploratory data analysis* (Sage University Papers on Quantitative Applications in the Social Sciences). Beverly Hills, CA: Sage.

Mosteller, F., & Tukey, J. W. (1977). *Data analysis and regression: A second course in statistics.* Reading, MA: Addison-Wesley.

NIST/SEMATECH. (2003). *E-handbook of statistical methods.* Retrieved from http://www.itl.nist.gov/div898/handbook/eda.

Tukey, J. W. (1977). *Exploratory data analysis.* Reading, MA: Addison-Wesley.


# 探索性因子分析（Exploratory Factor Analysis）

参见因子分析【Factor Analysis】。

# 外在效度（External Validity）

外在效度是一种使研究者能将研究的发现推广到一个较大的总体【Population】的性质。在构建实验和非实验的（观察的【Observational】）研究设计时，就会涉及外在效度问题。研究者可通过小心地构建自己的研究设计来确保外在效度。

外在效度的威胁来自三个方面：无代表性样本【Sample】、人造的实验室环境和检验效应。研究的主要目的之一是将发现推广到更大的总体。我们来看一个有关负面的竞选广告对选民投现任总统候选人的票的影响的研究。为了使这一个研究可行，我们不可能对所有的选民，即感兴趣的总体进行调查。相反，我们应该对总体的一个样本进行研究。我们需要特别予以关注的是，样本和感兴趣的总体的潜在的因果过程必须是相同的。有时，找到一个代表样本【Representative Sample】是很困难的。许多时候，参与者是从当地的社区或研究进行的大学的校园招聘来的。一个基于大学低年级学生的样本的竞选广告，得到的结果不可能类似于从感兴趣总体得到的结果。低年级大学生可能更容易受广告的影响，因为他们还没有形成很强的党派联系，然而年长选民受负面广告的影响就比较小。

研究进行的环境设置可能会对发现有影响。实验常常都在实验室环境中研究现象。这种非自然的环境使身处其中的参与者知道自己正在被研究，因而有可能产生意外的结果。在负面的广告研究中，参与者可能先被要求观看各种广告，然后测量他们有关现任候选人的观点。然而，我们知道大多数人都不会主动地去注意电视上的商业广告，在电视节目因播出广告而中断时，许多人便一起聊天，有的人去拿小吃，有的人则干脆换台。因此，那种在人工环境中观看的广告，可能会使参与者对现任候选人给出差评，而在家观看同样的商业广告却没有影

响。如果观察发生在相对于自然环境的实验室环境,观察研究也可能会因为人造的环境而遇到问题。

对外在效度的最后一个威胁是测试的前测【Pretest】和已被测试的经验可能会改变治疗效应的大小。为了能测量负面广告对现任候选人赞成的影响效应,在参与者看到这个广告之前,我们可能需要某种赞成的量度。一个前测可能会使参与者变得更善于观察,或对现任候选人更有成见,从而产生令人误解的结果。在观察研究中,研究者必须确保观察者的存在不会影响应试者的行为。

第一,研究者在进行研究设计时,都会通过考虑这些威胁来提高外在效度。在探索电视上的晚间新闻的议题设置效力的研究中(agenda-setting power),延加、彼得斯和金德(Iyengar, Peters,& Kinder,1982)阐述了如何可以做到这一点。他们操控着在晚间新闻报道中报道的新闻量,并观察这是否会对参与者如何看这个问题有影响。

延加、彼得斯和金德的样本来自美国康涅狄格州纽黑文市的志愿者。这就有可能产生一个问题,因为纽黑文市代表不了美国的人口总体(感兴趣的人口总体)。为了确保他们的结果能推广到感兴趣的人口总体,他们尝试通过一个分类广告来征集参与者,力求参与者能多样性。提高样本的异质性可以使研究更具有概括性(Frankfort-Nachmias & Nachmias, 2000)。他们还报告了参与者的人口组成,允许读者自行判断实验的普遍性。他们也在他们的控制组和实验组之间采用了随机分配【Random Assignment】,以确保这些组之间不存在系统差异。

第二,研究者采取了几个步骤来防止实验过程产生假的发现。参与者结成小组来观看新闻报道,他们也从来未被鼓励对报道要予以特别的注意。实际上,研究者报告说,许多参与者在报道期间都在忙于交谈。这表明尽管实验并未在参与者的起居室(感兴趣的环境),但研究者却成功地复制了一个类似的环境,包括干扰。还应对测试环境的地点予以周密的考虑,以使它尽可能地类似于自然的环境。另一种选择是使用随机化的田野实验【Field Experiment】(Gerber & Green, 2000)。

研究者还应采取几个预防措施,以防止来自检验的假结果。这样的措施包括给参与者递呈一封有说服力的表面故事【Cover Story】,并用一种无引导方式(nonleading way)来设计预测。幌子可能是一种有用的工具,但是研究者也必须考虑到因为使用幌子欺骗参与者而带来的伦理问题。正是通过对效度威胁的周密思考,延加等人才能构建出将这些威胁的影响减少到最小的研究设计。

法兰克福-纳西米亚思和纳西米亚思(Frankfort-Nachmias & Nachmias,2000)认为在构建研究设计时,研究者必须意识到,他们必须在外在和内在效度【Internal Validity】之间作出权衡。内在效度确保自变量【Independent Variable】能使因变量【Dependent Variabl】产生变化。一个有着很强的外在效度,但却有着较弱的内在效度的实验,对促进我们的理解贡献很小,因为我们无法从正在研究的关系【Relationship】中得出结论。为了确保内在效度,很多时候,我们必须牺牲外在效度。在延加等人(Iyengar, Peters,& Kinder, 1982)有关议题设定的实验中,他们让参与者在自己的起居室内看新闻,从而给他们提供了更为自然的环境。然而,这还会妨碍研究者对参与者观看节目的控制,于是,为了保证内在效度,他们牺牲了一定的外在效度,将实验置于一个可控的环境中。在构建一个研究设计时我们必须知道,选择会同时影响外在和内在效度两者。

——Heather L. Ondercin

(沈崇麟译校)

## 参考文献


Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Chicago: Rand McNally.

Frankfort-Nachmias, C., & Nachmias, D. (2000). *Research methods in the social sciences* (6th ed.). New York: Worth.

Gerber, A. S., & Green, D. P. (2000). The effects of personal canvassing, telephone calls, and direct mail on voter turnout: A field experiment. *The American Political Science Review*, *94*, 653-664.

Iyengar, S., Peters, M. E., & Kinder, D. R. (1982). Experimental demonstrations of the "not-so-minimal" consequences of television news programs. *The American Journal of Political Science*, *4*, 848-858.


# 外推法(Extrapolation)

外推法也称为趋势外推法或曲线拟合，是一种预测技术。分析员依据过去时段的数据绘制数据点，选出此数据的一条最佳拟合趋势线（或曲线），然后再延伸那条趋势线以投射未来的值。外推法的核心是双变量回归技术，其中自变量【Independent Variable】是时间，而需要投射的变量就是因变量【Dependent Variable】。

史密斯、塔伊曼和斯旺森（Smith, Tayman, & Swanson, 2001）强调，外推法技术最典型的特征就是历史值只能用于投射被研究的变量未来的值。外推法技术有许多变化形式，从最简单的线性回归到更复杂的多项式和增长曲线（logistic curves），有些变化形式也把比率也纳入了投射模型。

人口学家和规划人员通常使用外推法来推论总体的数字，公共管理者也使用外推法来进行项目预算和开支预测，商务人士也用外推法来预测销售和收益。外推法之所以备受青睐，主要是因为简单，需要的数据也不多。使用电子制表程序或计算器，就可以很容易地生成投射使用的外推技术。就数据的要求而言，最少仅需要两个数据点就可以形成一条投射未来值的趋势线。虽然外推法在运用中相对简单，但史密斯（Smith, 2001）报告了许多经验研究，证明它与其他复杂方法，例如同期群成分技术（the cohort-component technique）一样准确。

图 1 展示了使用外推法来为一个中等县投射人口的案例，使用了三种变化形式：一条线性曲线、一条几何曲线和一条抛物线。依据美国人口调查局（U. S. Census Bureau）的数据，1940—2000 年的总人口数字被用来投射这个县 2010、2020、2030 年的人口。图 1 显示的人口投射的范围显然与所选的曲线有关。从表面来看，抛物线似乎是这种情况下"拟合最好"的曲线。

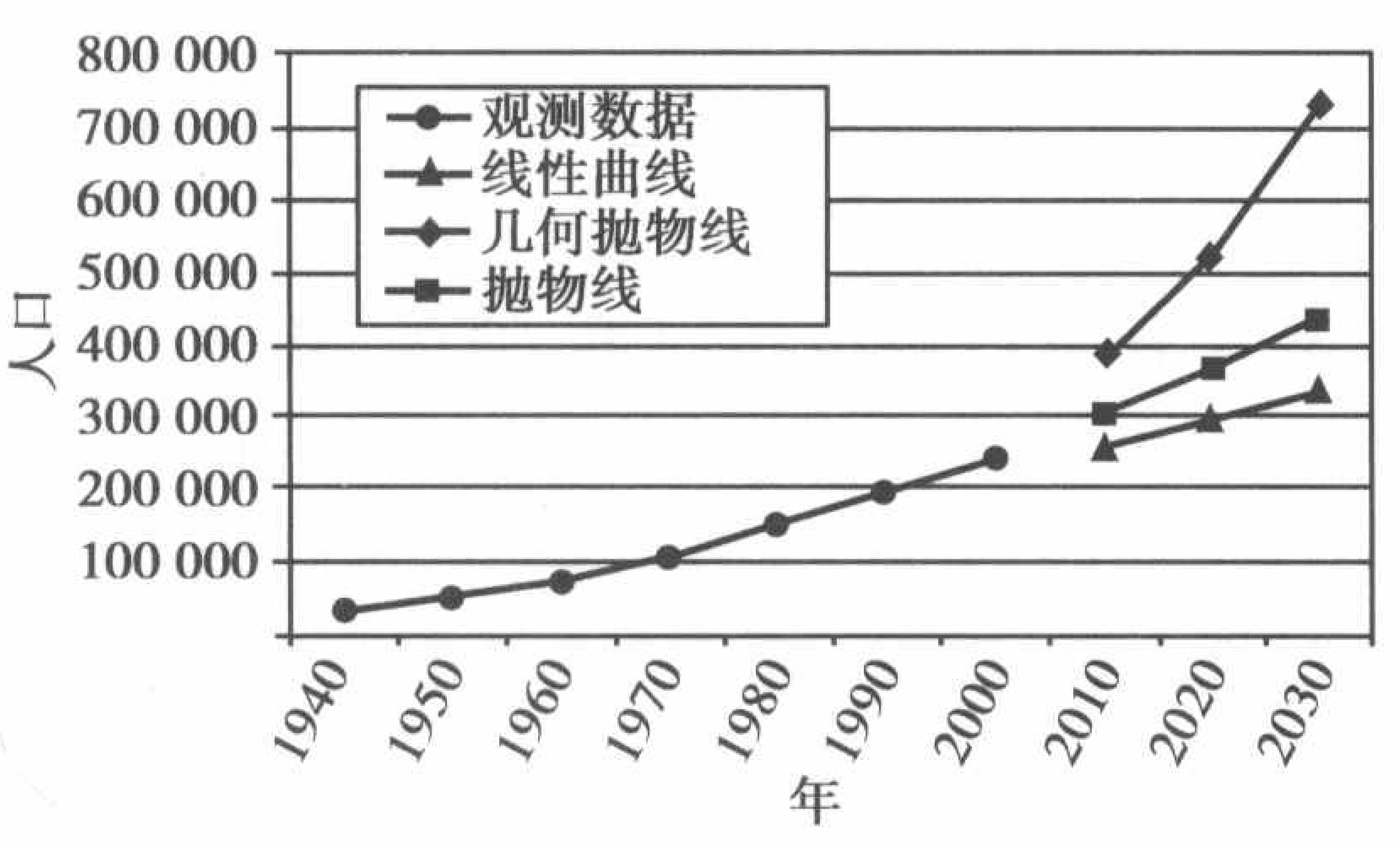



**图 1　外推技术的应用例解**

外推法简单、对数据要求不高，使它成了一种吸引人的投射方法，但也成了这项技术的主要局限。外推法依赖聚合数据输入，生成的也是聚合的结果。那些可能会影响研究变量的因素也是外在于这种方法的。比如说，在进行人口投射时，外推法技术不考虑住房趋势效应、经济变化或其他人口的

外在压力。输入模型的数据很简单,输出的也很简单。比如说,当采用外推法来投射政府收入时,这种方法只会产生一个总收入;不会提供来源于各种资金的收入变化。

这个模型的另一个局限是,外推法用过去的情况来投射将来的情况。尽管过去的情况是预测未来的有用的起始点,但没什么能保证过去的趋势会持续下去。最后,时间段的选择和分析中使用的数据的观察数也可能会对结果有极大的影响。有鉴于此,外推法的使用应十分谨慎,对它在投射未来时的效用和局限性要有充分的了解。

——Tim Chapin

(杨可译　沈崇麟校)

## 参考文献

Smith, S. K., Tayman, J., & Swanson, D. A. (2001). *State and local population projections: Methodology and analysis.* New York: Kluwer Academic/Plenum Publishers.

## 极端个案(Extreme Case)

极端个案就是在某个变量上的值达到了可能取值的极限的个案【Case】。比如说,在研究某个社区的年龄时,最极端的上限值可能是 95 岁,而最极端的下限值可能是 18 岁。这些代表了极端个案,但他们可能是也可能不是离群值【Outliers】。离群值必须远离其他的值,但极端个案不一定如此。比如说在年龄研究中,可能有很多被调查者都是 80 多岁、90 多岁,这可能意味着有 95 岁的,尽管很极端,但并没有远离邻近的值。相比之下,如果 95 岁以下最高的年龄是 72 岁,那么显然 95 岁就是一个离群值,同时也是一个极端个案。

——Michael S. Lewis-Beck

(杨可译　沈崇麟校)

# F

## $F$ 分布($F$ Distribution)

$F$ 分布也称为方差率分布,是一种从估计同一总体的方差【Variances】($\sigma^2$)的两个样本方差($s^2$)的比率导出的抽样分布【Samplimg Distribution】。这个分布原由罗纳德·费希尔(Fisher,1890—1962)提出,后经斯内德克(Snedecor,1881—1974)所发展,他将它命名为 $F$ 分布,以表示对费希尔的敬意。

### 用 $F$ 分布进行显著性检验

与费希尔和斯内德克对它的开发目的一致,$F$ 分布通常的使用都与方差分析【Analysis of Variance】连在一起。然而在最一般的显著性检验【Significance Testing】应用中,$F$ 分布在检验两个总体的方差相等的假设中十分有用:

$$H_0:\sigma_1^2=\sigma_2^2$$
$$H_1:\sigma_1^2\neq\sigma_2^2$$

为了理解 $F$ 分布和它在检验这样的假设中的使用,让我们来设想 $H_0$ 为真。这就是说,考虑两个可能有着不同的均值($\mu$),但却有着相同的方差($\sigma^2$)的正态分布的总体。如果我们从一个总体取一个样本量为 $n_1$ 的随机样本,而从另一个总体取一个样本量为 $n_2$ 的随机样本,那么我们可以通过构建样本方差,估计来自每一个样本的总体的方差。

$$\hat{\sigma}^2=s^2=\frac{\sum(X-\bar{X})^2}{n-1}$$

样本方差的分母($n-1$)是与那一统计量有关联的自由度【Degrees of Freedom】,我们可以把它标为 $v$。即使两个总体有相同的方差,抽样误差【Sampling Error】也可能会导致不同的样本方差。然后,我们可以构建两个样本方差的比率,它便是 $F$ 比率【$F$ Ratio】:

$$F=\frac{s_1^2}{s_2^2}=\frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2}$$

用 $H_0$ 为真,我们可以通过构建样本量为 $n_1$ 和 $n_2$ 的每一可能组合来导出 $F$ 比率的抽样分布,它将导致 $v_1$(即 $n_1-1$)和 $v_2$(即 $n_2-1$)的抽样分布(用 $H_0$ 为假,分布称为无中心 $F$ 分布[noncentral $F$ distribution]。它不容易设定,但它却比 $F$ 分布更对称)。

作为一个 $F$ 分布是怎样用于检验两个总体方差相等的假设的例子,我们来考虑一下两个样本量分别为 $n_1=31$ ($v_1=30$)和 $n_2=41$($v_2=40$)的样本。因为我们最感兴趣的常常是分布的尾部,那里都是不典型的值,统计教科书一般都含有一张 $F$ 比率表,它确定了 $F$ 分布的上端的百分比(上端2.5%,上端5%等)。如果我们采用了0.05

的 $\alpha$ 水平，那么 $F$ 比率将切去有 $v_1 = 30$ 和 $v_2 = 40$ 的 $F$ 分布上端的 5%（称为 $F_{临界值}$），等于 1.74。因此，根据 $F$ 分布，在 $H_0$ 为真时，只有 5% 的 $F$ 比率等于或大于 1.74（同样，统计程序将会产生一个 $H_0$ 为真时的概率值）。在掌握了这种信息之后，我们就可以检验两个样本方差取自有相同方差的总体的虚无假设。如果我们两个样本的 $F$ 比率值等于或大于 1.74（或概率小于我们的 $\alpha$ 水平，一般为 0.05），我们则倾向认为，抽取分子的样本的总体方差大于抽取分母样本的总体方差。如果 $F$ 比率大于 1.0 且小于 1.74，我们将保留 $H_0$（如果 $F_{临界点}$ 小于 1，只要将比率颠倒，并把它与有 40 和 30 自由度的 $F_{临界值}$ 相比较，在 $\alpha = 0.05$ 时，它等于 1.79）。

## $F$ 分布的特性

因为 $F$ 分布的形状受与两个样本方差有关的自由度（$v_1$ 和 $v_2$）的摆布，所以没有单一的 $F$ 分布，只有一个分布族。$F$ 分布中的 $F$ 比率总落在 0 到无限大之间。只要 $v_2$ 大于 2，$F$ 分布的模式便是

$$\left(\frac{v_1 - 2}{v_1}\right)\left(\frac{v_2}{v_2 + 2}\right)$$

只要 $v_2$ 大于 2，$F$ 分布的均值就等于

$$\frac{v_2}{v_2 + 2}$$

只要 $v_2$ 大于 4，$F$ 分布的方差就等于

$$\frac{2v_2^2(v_1 + v_2 - 2)}{v_1(v_2 - 2)^2(v_2 - 4)}$$

$F$ 分布一般都是单峰的和正偏态的（只要 $v_1$ 大于 2）。在正偏态分布中，均值将大于模式。例如在上述例子中，$v_1 = 30$ 和 $v_2 = 40$，模式 $F$ 比率为 0.89，而均值 $F$ 比率则是 0.95，与 $F$ 分布的正偏态一致。由公式可知，模式总是小于均值的，尽管随着 $v_1$ 接近无限大，这个差数将会变小。

$F$ 分布与使用很普遍的统计分布 $t$ 分布【*t*-Distribution】有关，亦即当 $v_1 = 1$ 时，$t^2 = F$。

——Hugh J. Foley

（沈崇麟译校）

＊也可参见分布【Distribution】。

## 参考文献

Hays, W. L. (1994). *Statistics* (5th ed.). Fort Worth, TX: Harcourt Brace.

Heyde, C. C., & Seneta, E. (Eds.). (2001). *Statisticians of the centuries*. New York: Springer.

Keppel, G. (1991). *Design and analysis: A researcher's handbook*. Englewood Cliffs, NJ: Prentice Hall.

Stuart, A., & Ord, J. K. (1994). *Kendall's advanced theory of statistics* (Vol. 1). London: Edward Arnold.

# *F* 比率（*F* Ratio）

$F$ 比率是方差或方差类数量的比率。一个简单的例子是在方差分析【Analysis of Variance】中检验组间的方差是否显著大于组内的方差。隐含于诸如这样的检验中的一般概念是组内个体间的差异都只不过是随机的差异而已（误差的方差），而组间均值的方差的差异则不是偶然的（备择假设）。

## 实例

举一个方差分解的例子来说明 $F$ 比率的计算方法。将有均值 5 和方差 54 的一系列数字 2 2 3 4 4 5 6 7 8 9 分解成两个组。我们可以设想对这两个组(如一组男人和一组女人)的某一定量变量的分数作一番考察,我们假定这个变量是以每天抽烟支数量度的抽烟行为。男人组的分数为 2 2 3 4 4,有均值 3 和方差 4。女人组的分数为 5 6 7 8 9,有均值 7 和方差 10。组内的变差为 4+10=14。组间的变差为组均值 3 和均值 7 所定义。它们的均值是 5,方差是 8。组间的变差完全由组确定,可用组之间的变差乘以每个组的个体数求得:8×5=40。我们不难看出 54=40+14 或 SST=SSB+SSW(总平方和=组间平方和+组内平方和)。我们对自由度进行类似的分解,将总和分解成组间和组内两个组成部分。计算系列数的均值失去了一个自由度。因此,10 个分数总起来共有 9 个自由度。对于组间的组成部分,我们计算了 2 个组,所以有 1 个自由度,对组内组成部分,我们对 2×5 个体进行了计算,每次计算组均值都失去了一个自由度,这样还剩下 2×(5−1)= 8 个自由度。我们在这里也看到了总和=组间+组内,因为 9=1+8。假如现在我们想使用方差,而不是变差(均方 MS,而不是平方和 SS),那么我们必须将每个有关的变差除以相应的自由度:MSB = 40/1 和 MSW = 14/8。组间的方差 MSB 和组内方差 MSW 的比率是一个统计量,$F=(40/1)/(14/8)=22.86$。这个统计量的抽样分布是分子的自由度为 1,分母的自由度为 8 的 $F$ 分布(之所以选择 $F$ 作为该分布的名称,是为了纪念罗纳德·费希尔对此做出的贡献)分子 1 个自由度和分母 8 个自由度的 $F$ 的临界值可以在理论的 $F$ 分布统计表上找到,对于公设的 0.05 的 $\alpha$,它等于 5.32。在我们的经验例子中求得的 $F$ 比率的值则大于它,因此即使只用这些比较小的数目,我们也有 95%的概率拒绝虚无假设。虚无假设表明比率 MSB/MSW 等于 1。换言之,组间的差异(MSB)并不大于随机假定的组内的个体之间的差异(MSW=误差的方差)。

## 在多变量分析中的许多应用

在多变量分析中,$F$ 比率常用于比上面的例子复杂得多的情境。方差分析中的组间差异检验,多元回归分析中的回归【Regression】模型的显著性检验和判别分析【Discriminant Analysis】中的中心差异检验等只不过是 $F$ 比率的许多应用中少数几个而已。拉奥对 $F$ 分布情有独钟,下大力开发了许多应用于多变量分析的 $F$ 变体。

## $F$ 分布和其他分布之间的关系

有心的读者已经发现,我们也可十分方便地使用均值差的 $t$ 检验来替代这个例子中的 $F$ 检验。我们可以从有标准误差 = $\left(\frac{s_w^2}{n_1}+\frac{s_w^2}{n_2}\right)^{1/2}$ 的均值差的抽样分布求得第一组的均值(3)和第二组均值(7)的差为(−4),而在这里 $s_w^2$ 等于$(4+10)/[(5-1)+(5-1)]=1.75$,而 $n_1=n_2=5$。在我们的配对样本中发现的均值差的 $t$ 值为:$t=(4-0)/[(1.75/5)+(1.75/5)]^{1/2}=4.78$。由 $t$ 表可知,两组均值的差异是显著的。前面求得的 $F$ 值为 22.86(在舍入误差之内),在自由度为 1 时,等于求得的 $t$ 值 4.78 的平方,即 $F=t^2$。

我们还可以看一下在 $F$ 和 $t$ 表中,1 个自由度分子的 $F$ 临界值(只有两个组!)等于对应的 $t$ 值的平方。

$F$ 分布【$F$-Distribution】属于 $\beta$ 分布的一个子族,有时也称为反向贝塔分布(inverted beta distributions)。它也和 $t$ 分布、卡方分

布、正态分布【Normal Distribution】和其他分布有联系。进一步的讨论请参见 F 分布【F-Distribution】。

——Jacques Tacq

（沈崇麟译校）

## 参考文献

Blalock, H. M. (1972). *Social statistics*. Tokyo: McGraw-Hill Kogakusha.

Fisher, R. A. (1953). *The design of experiments*. Edinburgh, UK: Oliver and Boyd.

Hays, W. L. (1972). *Statistics for the social sciences*. New York: Holt International.

Tacq, J. J. A. (1997). *Multivariate analysis techniques in social science research: From problem to analysis*. London: Sage.

# 表面效度(Face Validity)

表面效度是一个量度在多大程度上清楚而明白地刻画了它意图评估的建构估计值。因此,表面效度指的是一个测验的"易见性"——测验的目的在多大程度上对那些采用它的人是显而易见的。如果测验的内在目的是明确的,即使对那些幼稚的回答人也是如此,那么我们就说,测验有高表面效度;反之,如果测验的内在目的是不明确的,那么测验便有低表面效度(Nevo, 1985)。表面效度的概念与题项敏锐性类似,但是它们也有着重要的差别。题项的敏锐性描述的是单个测验题项的透明性,而表面效度描述的则是整个测验的透明性(Bornstein, Rossner, Hill & Stepanian, 1994)。我们有可能构建一个单个的测验题项的内在目的是不清楚的,但在把这些题项作为一个整体详察时,测验的目的就在总体上变得清晰了。

不同类型的测验表面效度具有的效应不尽相同。研究表明,高表面效度可能有利于智力、天资和成效测验。在测验的目的更为明确时,被测者便会不那么焦虑,因而会更加主动地接受测验,即使测验的题项具有很高的挑战性(Messick, 1995; Nevo, 1985)。

在一个测验被设计用来评估个性或精神病理学的某个方面时,高表面效度可能会适得其反。在这种情况中,高表面效度可能会使回答人的回答朝向他们认为的测验者想要看到的方向的偏倚。研究表明,幼稚的回答人在进行高表面效度的测验时,比那些进行低表面效度的测验的人更加容易"假好"和"假坏"(Bornstein, 2002; Bornstein et al., 1994)。

有两种一般的估计测验的表面效度的方法。有些心理计量学家一直采用一种方法,要求参与者确认来自一批可能选项的一个评估工具的目的。如果参与者能精确地做到这一点,那么测验便是有高表面效度的。另一些研究者,一直采用一种间接的方法,要求参与者故意弄偏他们的回答,提高或降低他们的测验分数。参与者能以这种方式在一定程度上更改他们的分数,那么测验便被认为是有高表面效度的,因而它对活体内自我展示的效应是敏感的。

考虑到表面效度对心理学测验的重要性,在测验开发期间,我们总是应当对表面效度加以评估和控制。有几种有用的策略。例如,测验题项的措辞可以更为精妙。这样测量的目的便可被遮盖起来。此外,可以对测验题项加以排列,这样那些来自一个分量表的题项就不会彼此出现在接近的地方,已经证明,这种做法也可以降低测验的表面效度。最后,可以将一个与测验的真实目的无

关的干扰题项包括进来,这样真正的测验题项的内容就会变得不那么明显。这后一种策略,在将两种方法合起来使用时,对于测验目的的掩饰特别有效。

——Robert F. Bornstein

(沈崇麟译校)

## 参考文献


Bornstein, R. F. (2002). A process dissociation approach to objective-projective test score interrelationships. *Journal of Personality Assessment*, *78*, 47-68.

Bornstein, R. F., Rossner, S. C., Hill, E. L. & Stepanian, M. L.(1994). Face validity and fakability of objective and projective measures of dependency. *Journal of Personality Assessment*, *63*, 363-386.

Messick, S.(1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, *50*, 741-749.

Nevo, B.(1985). Face validity revisited. *Journal of Educational Measurement*, *22*, 287-293.


# 因子分析(Factor Analysis)

一般来讲,因子分析一词是指许多相似但又有区别的多变量统计模型中的一种,这种模型将它的观察变量作为不可直接观察的一组潜变量【Latent Variable】或假设的变量,即所谓的因子的线性函数。

因子分析模型与回归【Regression】模型类似,它们都有作为自变量【Independent Variable】的线性函数的因变量【Dependent Variable】。但是与回归不同,因子分析的自变量不是独立于因变量观察的。

根据因子变量是确定的还是不确定的,因子分析模型还可以作进一步的区分。如果它们接着能被导出为观察变量的线性函数,那么因子就是确定的。否则就是不确定的。确定模型包含着各种成分分析模型,如主成分分析【Principal Components Analysis】(Hotelling, 1933; Joliffe, 1986; Pearson, 1901);加权主成分分析(Mulaik, 1972)和哥特曼(Guttman,1953)图像分析。不确定模型为公共因子模型所代表(Spearman, 1904; Thurstone, 1947),它力图对观察变量之间的协变作出解释。在这里,协变是观察变量在不同程度上分享一组公共因子变量的变差影响的结果。

确定因子分析常用于简化数据,方法是找出少数几个能捕捉到观察变量中的变差和协变信息的变量。确定因子上的分数可以用观察变量的线性组合来计算。这些因子分数在其他的多变量统计程序中,如在多元回归或多元方差分析中可以用作自变量和因变量——就像股权认证一般。然而从理论上讲,这些确定的成分分析的主要不足在于,它们的因子是一种统计上的人工产品,对于确定它们的那组观察变量是唯一的(Mulaik, 1987)。变化观察变量,我们就会得到不同的线性组合。成分因子离开那组线性组合的观察变量便无法独立存在。

相反,公共因子模型的公共因子对观察变量是不确定的,也不是它们的线性组合。公共因子可与在公共因子分析的简单理论结构中那些独立存在的变量对应(Thurstone, 1947);来自某一领域的不同的观察变量的集合可以是那些相同的公共因子的线性函数。因此,就发现那些作为其他变量的共同原因的自主变量这一目的而言,它具有理论的重要性,公共因子模型一般都

优于确定的成分分析模型。

然而公共因子模型在应用上会受到一定的限制。在观察变量无自然的次序可觅时,公共因子模型的使用便会受到限制。公共因子模型唯一允许的排序原则是在潜变量上的观察变量函数的相依关系。潜变量之间的函数相依的公共因子模型是无法可依的。潜变量之间的关系只可取相关或协方差的形式,它们是无方向和无次序的。但是那些在时间、空间或某种属性的程度上有着自然次序的变量,可能与它们紧邻的那些变量更相关,而与在自然次序上远离它们的变量则较少相关。在这些情况中,除潜变量上的观察变量的函数相依关系外,每个依次对应于观察变量中的某一个连续的潜变量也许仅仅是在次序上紧挨它前面的潜变量和它所特有的某些新的潜变量的线性函数。有这种性质的模型称为单形模型(Jöreskog, 1979)。处理这种情况和其他一些情况的更为恰当的方法是结构方程模型【Structural Equation Modeling】,因为它能在潜变量之间建立线性函数。因此,如果我们要恰当地将公共因子模型应用于一组观察变量,那么我们就必须在开始就确定变量之间不存在明显的自然的次序。尽管公共因子模型可能与这样一种有少量公共因子的数据拟合得很好,但是通常这些因子都在理论上难以解释(Jöreskog,1959)。

因子分析还可以区分为探索性因子分析和验证性因子分析。探索性公共因子分析主要是分析性的,用于我们认为对那些符合公共因子模型的某一变量域的成分因子变量知之甚少这样的场合。其目的在于分析观察变量,以"发现"更为基本的成分变量。验证性因子分析主要是进行综合,用于我们有足够的知识就作为几个变量共有的理论上的变量的函数建立假设的场合。该理论允许我们在某些约束条件内,预测观察变量间的协方差模式,并对这种模式进行检验(参见验证性因子分析【Confirmatory Factor Analysis】)。本词条的其余部分将介绍探索性公共因子分析。

## 探索性因子分析的方程

探索性因子分析的方程通常都用矩阵代数来表示。因此,公共因子分析的模型方程由以下矩阵方程给出:

$$\mathbf{Y} = \mathbf{\Lambda X} + \mathbf{E}$$

式中的 $\mathbf{Y}$ 是一个 $p$ 个观察的随机变量的 $p\times 1$ 列向量;$\mathbf{\Lambda}$ 是因子模式负荷 $\lambda_{ij}$ 的 $p\times m$ 矩阵,它类似于回归系数;$\mathbf{X}$ 是一个公共因子随机变量的 $m\times 1$ 列向量;而 $\mathbf{E}$ 则是一个 $p$ 个独特因子随机变量的 $p\times 1$ 列向量。

随后,我们假定:($a$)公共和独特因子变量是不相关的,即 $\text{cov}(\mathbf{X},\mathbf{E})=0$;($b$)独特因子是互不相关的,即 $\text{cov}(\mathbf{E})=\mathbf{\Theta}^2=$ 一个对角矩阵。

根据模型的方程和这些假定,我们可导出公共因子模型的基本定理,它可用下式表示:

$$\text{cov}(\mathbf{Y})=\sum=\mathbf{\Lambda\Phi\Lambda}'+\mathbf{\Theta}^2$$

式中的 $\sum$ 是 $p$ 个观察变量的 $p\times p$ 的方差-协方差矩阵;$\mathbf{\Lambda}$ 是元素 $\lambda_{ij}$ 的 $p\times m$ 的因子模式矩阵;$\mathbf{\Phi}$ 是公共因子之间的 $m\times m$ 的方差【Variances】和协方差【Covariances】矩阵;而 $\mathbf{\Theta}^2$ 则是独特因子变量的 $p\times p$ 的对角矩阵。因为 $\mathbf{\Theta}^2$ 是一个对角线之外的元素均为零的对角矩阵,所以协方差矩阵的 $\sum$ 的对角线之外的元素只取决于 $\mathbf{\Lambda\Phi\Lambda}'$ 的对角线元素,它是唯一的公共因子变量的函数。

## 进行因子分析的步骤：一个实例

步骤1：挑选或构建变量

最好以系统方式选择或构建一些代表某一活动领域的变量。瑟斯通（Thurstone，1947），因子分析领域的先驱者认为，生成一个域的公共因子定义了变量域。因此，我们必须先考虑那些存在于这一域内的可能的公共因子。然后，对于这些预期的因子中的每一个，我们应当选择或构建至少四个，我们认为是这些因子的比较纯净的量度的指标变量。一个预期因子的四个指标用多种因素确定该因子。

作为一个例子，卡尔森和穆莱克（Carlson & Mulaik，1993）选择了如表1所列的15个基于两极的个性等级量表的变量。他们期望这些变量代表四个个性因子：（a）友好（friendliness），由两极等级量表友好-不友好（friendly-unfriendly）、有同情心-无同情心（sympathetic-unsympathetic）、善良-不善良（kind-unkind）和有情-无情（affectionate-unaffectionate）代表。（b）能力（ability），由量表聪明-不聪明（intelligent-unintelligent）、能干-不能干（capable-incapable）、有能耐-没有能耐（competent-incompetent）和机灵-愚笨（smart-stupid）所代表。（c）外向（extraversion），由爱说话-不爱说话（talkative-untalkative）、开朗-孤僻（outgoing-withdrawn）、爱社交-爱独处（gregarious-solitary）、思想外向-思想内向（extraverted-introverted）和爱交际-不爱交际（sociable-unsociable）代表。他们预料诸如帮助的-合作和合作-不合作这样的变量，可能代表了某种友好和能力组合的概念，因而可能会有在这些因子上，再加上在（d）的不同于它们自身的其他因子上的负载。

步骤2：求变量的分数和计算相关矩阵

然后卡尔森和穆莱克（Carlson & Mulaik）让280名学生在由15个题项构成的等级量表上打分。打分完成后，计算了15个量表间的相关，将计算结果列在了表2中。

表1 用于生成因子分析的15个变量的15种两极化个性的等级量表

| | |
|---|---|
| 1. | 友好：_1_：_2_：_3_：_4_：_5_：_6_：_7_：不友好 |
| 2. | 有同情心：_1_：_2_：_3_：_4_：_5_：_6_：_7_：无同情心 |
| 3. | 善良：_1_：_2_：_3_：_4_：_5_：_6_：_7_：不善良 |
| 4. | 有情：_1_：_2_：_3_：_4_：_5_：_6_：_7_：无情 |
| 5. | 聪明：_1_：_2_：_3_：_4_：_5_：_6_：_7_：不聪明 |
| 6. | 能干：_1_：_2_：_3_：_4_：_5_：_6_：_7_：不能干 |
| 7. | 有能耐：_1_：_2_：_3_：_4_：_5_：_6_：_7_：没有能耐 |
| 8. | 机灵：_1_：_2_：_3_：_4_：_5_：_6_：_7_：愚笨 |
| 9. | 爱说话：_1_：_2_：_3_：_4_：_5_：_6_：_7_：不爱说话 |
| 10. | 开朗：_1_：_2_：_3_：_4_：_5_：_6_：_7_：孤僻 |
| 11. | 爱社交：_1_：_2_：_3_：_4_：_5_：_6_：_7_：爱独处 |
| 12. | 思想外向：_1_：_2_：_3_：_4_：_5_：_6_：_7_：思想内向 |
| 13. | 帮助：_1_：_2_：_3_：_4_：_5_：_6_：_7_：不帮助 |
| 14. | 合作：_1_：_2_：_3_：_4_：_5_：_6_：_7_：不合作 |
| 15. | 爱交际：_1_：_2_：_3_：_4_：_5_：_6_：_7_：不爱交际 |

表 2 15 个变量的相关

对称矩阵的下半部分

| | | 1 友好 | 2 有同情心 | 3 善良 | 4 有情 | 5 无情 | 6 能干 | 7 有能耐 | 8 聪明 | 9 爱说话 | 10 开朗 | 11 爱社交 | 12 外向 | 13 帮助 | 14 合作 | 15 爱交际 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 友好 | 1.000 | | | | | | | | | | | | | | |
| 2 | 有同情心 | 0.777 | 1.000 | | | | | | | | | | | | | |
| 3 | 善良 | 0.809 | 0.869 | 1.000 | | | | | | | | | | | | |
| 4 | 有情 | 0.745 | 0.833 | 0.835 | 1.000 | | | | | | | | | | | |
| 5 | 无情 | 0.176 | 0.123 | 0.123 | 0.112 | 1.000 | | | | | | | | | | |
| 6 | 能干 | 0.234 | 0.159 | 0.205 | 0.183 | 0.791 | 1.000 | | | | | | | | | |
| 7 | 有能耐 | 0.243 | 0.155 | 0.187 | 0.186 | 0.815 | 0.865 | 1.000 | | | | | | | | |
| 8 | 聪明 | 0.234 | 0.190 | 0.238 | 0.215 | 0.818 | 0.841 | 0.815 | 1.000 | | | | | | | |
| 9 | 爱说话 | 0.433 | 0.319 | 0.321 | 0.435 | 0.174 | 0.209 | 0.239 | 0.258 | 1.000 | | | | | | |
| 10 | 开朗 | 0.473 | 0.480 | 0.410 | 0.527 | 0.220 | 0.274 | 0.269 | 0.261 | 0.744 | 1.000 | | | | | |
| 11 | 爱社交 | 0.433 | 0.438 | 0.406 | 0.526 | 0.188 | 0.227 | 0.242 | 0.228 | 0.711 | 0.853 | 1.000 | | | | |
| 12 | 外向 | 0.447 | 0.396 | 0.350 | 0.500 | 0.192 | 0.221 | 0.227 | 0.224 | 0.758 | 0.846 | 0.801 | 1.000 | | | |
| 13 | 帮助 | 0.649 | 0.693 | 0.697 | 0.694 | 0.283 | 0.344 | 0.370 | 0.365 | 0.443 | 0.552 | 0.514 | 0.473 | 1.000 | | |
| 14 | 合作 | 0.662 | 0.692 | 0.676 | 0.679 | 0.311 | 0.345 | 0.375 | 0.351 | 0.431 | 0.557 | 0.514 | 0.493 | 0.740 | 1.000 | |
| 15 | 爱交际 | 0.558 | 0.543 | 0.510 | 0.632 | 0.213 | 0.289 | 0.287 | 0.287 | 0.745 | 0.886 | 0.820 | 0.830 | 0.631 | 0.626 | 1.000 |

步骤 3:确定因子数

因子的抽取以因子之间的相关矩阵为根据。然而,这时我们必须确定用来求因子模式系数的估计值,公共因子的相关和独特的因子方差的因子抽取方法,它们将是对分析作最终解释的基础。用于这一分析的程序是用于 Macintosh 计算机(美国苹果公司生产的计算机) SPSS FACTOR Version 4.0(一种统计分析软件。在该软件的较新版本中,这一程序基本上没有什么变化)。

该程序为我们提供了几种适合抽取公共因子的方法。这些方法包括主轴抽取法(基本上是一种未加权最小二乘);未加权最小二乘(同一结果略有不同的算法【Algorithm】;广义最小二乘【Generalized Least Squares】法和最大似然【Maximum Likelihood】法。所有这些抽取法都需要先确定抽取的因子数,因为这一数字在整个迭代过程中必须固定在这一值上。SPSS 计算每一因子分析的主成分分析,然后列出修正的相关【Correlation】矩阵 $R$ 的特征值【Eigenvalues】。这些数值可用来估计公共因子数。因子数可用碎石图【Scree Plot】来确定。陡坡图是每一特征值比照它在特征值的递降系列中的次序位置的量级图。这一图上点的连接显示,第一个大的特征值大于其他的特征值,随后是那些虽然小一些,但仍然有较大的量级的特征值。不过通常都会有一个点,在那儿,特征值的量级急速下降,突然变得平缓,其余的特征值几乎呈线性下降。许多人都认为这个点的邻近之处,变量开始受公共因子的显著影响,因而这一数目被用作公共因子的抽取数。也有些人认为,因子数应由 $R$ 的特征值大于 1 的个数来确定。但是大家都知道,这会使一定数目的公共因子有一个最弱的下界,因而可能会有更多的较小影响的公共因子。

这一版本及随后版本的 SPSS FACTORS 模块都未能报告那些能在理论上揭示公共因子数的不同的特征值组。考虑上述因子分析的基本定理方程,我们可以用 $\mathbf{R}$ 替代 $\Sigma$ 来代表相关矩阵。哥特曼(Guttman, 1954, 1956)证明,确定数目的公共因子的强下界是矩阵 $\mathbf{R}-\mathbf{S}^2$ 的正特征值数,式中的 $\mathbf{S}^2=[\mathrm{diag}\mathbf{R}^{-1}]^{-1}$。$\mathbf{S}^2$ 是一个强上界近似 $\mathbf{\Theta}^2$。这一正特征值数常常位于$p/2$左右处。不仅如此,许多 $\mathbf{R}-\mathbf{S}^2$ 的特征值都相当小,几乎为零。因此,许多研究者喜欢考察特征值的图形,以决定特征值开始有实质意义的值点。雷斯括格(Jöreskog, 1967)将哈里斯(Harris, 1962)开发的方法应用于最大似然因子分析算法开发,建议我们可以用 $\mathbf{S}^{-1}$前乘或后乘 $\mathbf{R}-\mathbf{S}^2$,得到 $\mathbf{S}^{-1}(\mathbf{R}-\mathbf{S}^2)\mathbf{S}^{-1}=\mathbf{S}^{-1}\mathbf{R}\mathbf{S}^{-1}-\mathbf{I}$。这一转换并没有改变得到的矩阵的正特征值数,这样,这个矩阵的正特征值数便等于要保留的因子数。但是这个矩阵的特征值中的每一个都对应于矩阵 $\mathbf{S}^{-1}\mathbf{R}\mathbf{S}^{-1}$的一个特征值。这样$\mathbf{S}^{-1}\mathbf{R}\mathbf{S}^{-1}$的任何一个大于0.1 的特征值都对应于 $\mathbf{S}^{-1}\mathbf{R}\mathbf{S}^{-1}-\mathbf{I}$ 的一个特征值。因此,研究者也可以通过考察 $\mathbf{S}^{-1}\mathbf{R}\mathbf{S}^{-1}$特征值来确定保留的公共因子数。

单独计算这一例子的 $\mathbf{S}^{-1}\mathbf{R}\mathbf{S}^{-1}$特征值显示,15 个特征值中只有 7 个大于 1.00。这些特征值分别是 37.484, 14.467, 9.531, 1.436, 1.289, 1.116 和 1.065。尽管三个因子有着实质性的贡献,但是我们还是多取了一个,因为这个因子是理论上有所期待的,而那些在它之后的因子则只有小得多的贡献(如果我们把每一个特征值都减去 1 的话)。

步骤 4:未旋转模式矩阵的因子抽取

在这个例子中使用的因子抽取法是最大似然法。这种方法比较稳定,即使在变量不是多元正态分布时,它也能在带有微分出来的公共因子的变量中最大化偏相关行列式,不论分布的形式是什么。在这一例中,未旋转因子模式矩阵的估计值由 $\hat{\mathbf{\Lambda}}=\hat{\mathbf{\Theta}}\mathbf{A}_m[\gamma_i-1]_m^{1/2}$给出,式中的 $\hat{\mathbf{\Theta}}$ 是迭代估计

的对角独特方差矩阵的平方根，$\mathbf{A}_m$ 是 $p\times m$ 矩阵，它的列是 $\hat{\mathbf{\Theta}}^{-1}\mathbf{R}\hat{\mathbf{\Theta}}^{-1}$ 的前 $m$ 个特征向量，而$[\gamma_i-1]_m$ 是 $m$ 阶的对角矩阵，它的对角元素由 $\hat{\mathbf{\Theta}}^{-1}\mathbf{R}\hat{\mathbf{\Theta}}^{-1}$ 的前 $m$ 个最大的特征值 $\gamma_i$ 中的每一个减去 1 形成的。在这种解中，公共因子互不相关。但是旋转的因子模式矩阵只是一个中间解。它定义了含任何 $m$ 个公因子的最大公共方差的一个公共因子空间。它不用于因子的解释。

*步骤 5：因子旋转*

在数学上，因子模式矩阵不是唯一的。若模型方程为 $\mathbf{Y}=\mathbf{\Lambda X}+\mathbf{E}$，那么我们就可以通过对它们进行线性转换，找到备择的因子模式矩阵和公共因子，令 $\mathbf{T}$ 为一个 $m\times m$ 的满秩转换矩阵，而 $\mathbf{T}^{-1}$ 则是它的逆矩阵，那么 $\mathbf{Y}=\mathbf{\Lambda TT}^{-1}\mathbf{X}+\mathbf{E}=\mathbf{\Lambda}^*\mathbf{X}^*+\mathbf{E}$，有 $\mathbf{\Lambda}^*=\mathbf{\Lambda T}$ 和 $\mathbf{X}^*=\mathbf{T}^{-1}\mathbf{X}$。这同样也有因子分析方程式。随之而来的问题是，究竟哪一个因子模式矩阵才是我们的解呢？瑟斯通(Thurstone，1947)用简单结构理念解决了这一问题。他论证道，如果大多数观察变量并非它们域内的所有公共因子的函数，那么，在公共因子空间中代表它们的向量将落入“坐标超平面”或全公共因子空间的子空间中。每个超平面或子空间至多由 $m-1$ 个公共因子生成，而这些子空间内的所有变量将只是那些生成该超平面的公共因子的函数。然后，全公共因子空间的公共因子，将在这些“坐标超平面”的交集中发现，因为它们将在 $m-1$ 个超平面的每一个的基向量中。如果坐标超平面能在发现那些占据这些低维的子空间的变量的子空间时予以确认，那么，公共因子便可以在它们的交集中予以确认。不仅如此，该域的变量的不同选择，仍然能在超平面的交集确认相同的公共因子。因此，因子的解对该域变量的选择不是唯一的，而对该域的所有观察变量的选择才是客观不变的结果。为了发现这些超平面，瑟斯通建议用插入这些公共因子空间的称为参照轴的超平面向量。公共因子空间的一个子空间，将是一个与参照轴正交的向量子空间，因而就像一把用参照轴作为把手，伞骨代表子空间中的向量的伞。因此，我们可以在公共因子空间内，四处移动每一个参照轴来寻找落入与参照轴正交的“伞”的不同的变量集。这些变量集将定义那些坐标超平面。瑟斯通原来使用每个因子对的二维图，并在这些二维空间中手动地移动它的参照轴，来发现那些将成行与参照轴正交的变量集。这是一种进度缓慢的工作。后来随着计算机的出现，简单结构的分析标准的建立，而使发现简单结构的解的过程变成了自动进行的。而更为重要的是，我们应该认识到简单结构并不意味着不相关或正交的因子。这些公共因子可能是相关的，因此，如果我们想要真正的简单结构解，那么就应当避免诸如最大方差【Varimax】这样的迫使公共因子相互正交的旋转方法。常用的简单结构的上佳算法是直接 Oblimin 和 Promax 法。其他的算法，据说略优于这两种算法，但是在商业因子分析软件中，一般都不包含这些算法。

在我们的例子中，使用直接 Oblimin 旋转法，将四个因子解旋转为简单结构。程序生成了一个因子模式矩阵、一个因子结构矩阵和因子之间的相关矩阵。因子模式矩阵包含像回归权数那样的来自公共因子的观察变量的权数。这个矩阵在解释因子时最有用，因为这些权数在跨选择的应试者子总体的范围限制下，将是不变的。不仅如此，模式权数还清楚地显示，那些变量究竟是不是那些公共因子的函数。因子结构矩阵包含着观察变量和公共因子之间的相关。该矩阵是因子模式矩阵和因子相关矩阵的积，$\mathbf{\Lambda\Phi}$。因此，一旦这两个矩阵求得，因子结构矩阵便也迎刃而解。此外，因子结构矩阵和

因子之间的相关矩阵在范围限制或主题选择上并非一成不变的。

步骤 6:因子的解释

因子的解释方法是排除归纳法(eliminative induction)。我们从上向下看那些在因子上有着大的"负载"因子模式矩阵的每一列。我们把因子作为假设的变量来解释。假设的变量为那些有着大的负载的变量所共有,但却不为那些几乎为零的负载的变量所有。在表 3 中,我们发现有些东西对他人好像有正的取向,或"善良"是那些在第一个因子上有着大的负载的变量所共有。能力似乎是第二个因子的共有成分,而外向则是第三个因子的共有成分。

**表 3 因子模式负载**

| | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 友好 | 0.783 17 | 0.045 61 | 0.069 36 | 0.022 96 |
| 同情心 | 0.845 25 | −0.053 40 | −0.013 19 | 0.122 06 |
| 善良 | 1.030 68 | 0.023 53 | −0.091 71 | −0.044 96 |
| 有情 | 0.794 70 | −0.036 80 | 0.146 17 | 0.036 76 |
| 聪明 | −0.077 08 | 0.886 50 | −0.014 61 | 0.039 89 |
| 能干 | 0.024 54 | 0.925 86 | 0.012 84 | −0.029 59 |
| 能耐 | −0.071 17 | 0.903 17 | 0.003 33 | 0.101 89 |
| 机灵 | 0.107 94 | 0.911 76 | 0.006 14 | −0.090 64 |
| 爱说话 | 0.011 44 | 0.032 35 | 0.841 01 | −0.090 20 |
| 开朗 | −0.045 48 | −0.000 38 | 0.923 07 | 0.084 21 |
| 社交 | 0.015 53 | −0.014 68 | 0.880 10 | 0.010 90 |
| 外向 | −0.026 96 | −0.014 28 | 0.938 15 | −0.030 29 |
| 有帮助 | 0.328 64 | 0.114 50 | 0.127 44 | 0.449 96 |
| 合作 | 0.234 77 | 0.100 52 | 0.119 66 | 0.568 80 |
| 社交能力 | 0.078 60 | 0.005 42 | 0.824 72 | 0.117 74 |

第四个因子确实在一定程度上只为帮助和合作所共有,而这两个变量诚如我们所料,在善良和能力上也有一些负载。表 4 中的因子之间的相关指出,善良和外向共同分享着某些东西,而与能力分享的东西则比较少。在有些情况中,因子分析者将分析因子之间的相关,以得到二阶的公共和独特的因子。

**表 4 公共因子相关**

| | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 因子 1 | 1 | 1.000 00 | | |
| 因子 2 | 0.218 57 | 1.000 00 | | |
| 因子 3 | 0.523 71 | 0.282 26 | 1.000 00 | |
| 因子 4 | 0.729 44 | 0.337 31 | 0.524 43 | 1.000 00 |

——Stanley A. Mulaik
（沈崇麟译校）

## 参考文献


Carlson, M., & Mulaik, S. A.(1993). Trait ratings from descriptions of behavior as mediated by components of meaning. *Multivariate Behavioral Research*, *28*, 111-159.

Guttman, L.(1953). Image theory for the structure of quantitative variates. *Psychometrika*, *18*, 277-296.

Guttman, L. (1954). Some necessary conditions for commonfactor analysis. *Psychometrika*, *19*, 149-161.

Guttman, L. (1956). "Best possible" systematic estimates of communalities. *Psychometrika*, *21*, 273-285.

Harris, C. W. (1962). Some Rao-Guttman relationships. *Psychometrika*, *27*, 247-263.

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, *24*, 417-441, 498-520.

Joliffe, I. T. (1986). *Principal component analysis*. New York: Springer-Verlag.

Jones, M. B. (1959). *Simplex theory* (U. S. Naval School of Aviation Medicine Monograph Series No. 3). Pensacola, FL: U.S. Naval School of Aviation Medicine.

Jöreskog, K. G. (1967). Some contributions to maximum likelihood factor analysis. *Psychometrika*, *32*, 443-482.

Jöreskog, K. G. (1979). Statistical models and methods for analysis of longitudinal data. In K. G. Jöreskog & D. Sörbom (Eds.), *Advances in factor analysis and structural equation models*. Cambridge, MA: Abt.

Mulaik, S. A. (1972). *The foundations of factor analysis*. New York: McGraw-Hill.

Mulaik, S. A. (1987). A brief history of the philosophical foundations of exploratory factor analysis. *Multivariate Behavioral Research*, *22*, 267-305.

Pearson, K.(1901). On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*, *6*(2), 559-572.

Spearman, C. (1904). General intelligence, objectively determined and measured. *American Journal of Psychology*, *15*, 366-374.

Thurstone, L. L. (1947). *Multiple factor analysis*. Chicago: University of Chicago Press.


# 析因设计(Factorial Design)

析因设计涉及在某一个或某一些其他变量上的两个或更多个变量的同步研究。因子都是可以被操控或量度的。每一个因子都由两个或更多水平或类别组成。水平可以在质量或数量上有所不同。例如,质性的婚姻状态可以由五个水平组成:从未结过婚、已婚、分居、离婚和鳏寡;量性因子可以包含特定的值(如递增的毒品计量),或值域(如四个年龄组:20~29, 30~39, 40~49和50~59)。

析因设计还可用因子的数目、因子数和每一个因子内的水平数,个案对因子或治疗是否随机的和个案是否是在一个或多个场合被量度作进一步的描述。因子也可以称为"路径",这样,一个双因子的设计就可以称为双路设计,一个三因子的设计可以称为三路设计,如此等等。有时,这种设计也可以用每个因子内的水平数来描述。一个由三个因子、两个因子有两个水平、一个因子有四个水平组成的设计可看作 2 × 2 × 4("一个 2 乘 2 乘 4")的析因设计。在一个完全随机化的析因设计中,每一个个案都被随机分配【Randomly Assigned】给一个,且只随机地被分配给一个因子水平或格的组合。

在一个混合析因设计中,至少有一个其个案在一个以上的场合被量度的因子和一个其个案只在一个场合量度的因子。

罗纳德·费希尔(Fisher,1935)认为,较之单因设计,析因设计有四个优点:第一个优点是更有效率或经济,在这样的设计中,对于同样的精度或力度,它所需要的个案或观察更少。与单因设计相比较,双因子或双路设计只需单因设计所需的个案的一半,三因素设计只需三分之一,而四因素设计则只需四分之一,如此等等(Snedecor, 1937)。之所以会这样,是因为一个因子的值被跨其他因子的值所平均。第二个优点是它更为全面,析因设计允许考察两个或更多的变量之间的相互作用。第三个优点是它能使结果具有更大的普遍性【Generalizability】。在这样的设计中,一个因子在更为广阔的范围内被调查。第四个优点是如果该因子没有和一个或更多个可能对其未解释或误差的方差的某些方面作出解释的因子有实质性相互作用的话,那么这样的设计将会给我们提供一个因子更为灵敏或有力的检验(Stevens, 2001)。

——Duncan Cramer

(沈崇麟译校)

*也可参见实验【Experiment】。

## 参考文献


Fisher, R. A. (1935). *The design of experiments* (1st ed.). Edinburgh, UK: Oliver and Boyd.

Snedecor, G. W. (1937). *Statistical methods* (1st ed.). Ames: Iowa State University Press.

Stevens, J. (2001). *Applied multivariate statistics for the social sciences* (4th ed.). Mahwah, NJ: Lawrence Erlbaum.


# 析因调查法(Rossi 法)(Factorial Survey Method [Rossis's Method])

人类形成那些有关世界运行方式的思想;人类也形成那些有关世界应该运行方式的思想。这些积极和规范的思想可用方程来代表。而这些方程分别称为积极信仰方程(positive-belief equation)和规范判断方程(normative-judgment equation)。ROSSI 析因调查法使我们得以对头脑中的方程进行估计。

积极信仰和规范判断方程与两个更进一步的方程——一个描述信仰或判断的组成成分的决定因素方程(determinants equation)和一个描述信仰/判断结果的结果方程(consequences equation)相连。

例如,个人在形成有关收入决定因素思想的同时,也形成了作为自身和他人的收入的思想。有关实际收入的决定因素和仅仅只是收入的决定因素本身都是个人和社会因子的产物。诸如那些有关职业结构和宗教的信息也同样如此。不仅如此,有关实际收入的信息帮助人们做出许多决定和行为——学些什么、到哪里工作等,而仅仅有关收入本身的那些思想,则帮助人们做出许多其他的决定和行为,如投票、参加罢工、为院外游说组织作贡献和设计支付时间表等。

因此,积极信仰方程和规范判断方程的每一个都可以和决定因素方程合用,形成一个多水平的方程系统。同样,积极信仰方程和规范判断方程的每一个也都可以和结果方程合用,形成另一个(可能的多水平)方程系统。

析因调查法的先驱和开发者,彼得·罗西(Rossi,1951, 1979)和她的合作者(Jasso & Rossi, 1977; Rossi & Anderson, 1982;

Rossi & Berk, 1985; Rossi, Sampson, Bose, Jasso, & Passel, 1974)给我们提供了一个人们头脑中估计信仰/判断方程,检验回答人之间的同质性和估计决定因素和结果方程的整合框架。研究方案中的所有部分的设计都以用信仰/判断方程可以达到的最好的可能属性来求估计值为目的。

## 析因调查法中的数据收集

要求每一个回答人给特定的,与一个虚构的单位(比如说,一个人或家庭)对应的结果变量(例如健康状况、工资获得、公正的刑期)打分。这些单位用可能与之有关的特性,如年龄、性别、学习或饮食习惯、获得医疗或住房的难易程度等来描述。这样的描述称为"情景模拟"(情景模拟技术【Vignette Technique】)。罗西的一个重要观点是,先通过生成潜在相关的特性的所有水平的可能组合的总体【Population】,来达到对丰富多彩和复杂的现实的高度保真,然后再抽取回答人的随机样本。有时,情景模拟是用由许多可能的认识代表的许多特性来描述的,而这些特性是完全交叉的。罗西的设计还具有以下三个重要特征:(a)在模拟情景的总体中,模拟情景特性之间的相关为零或接近零,因此可降低或消除那些与多重共线性【Multicollinearity】有关的问题。(b)那些呈现给一个回答人的情景模拟都在调查者的控制之下(即它们都是固定的),这样只有在回答人没有给所有呈现给他们的情景模拟打分时,估计积极信仰和规范判断方程中的内生性问题才会出现。(c)一组大的情景模拟呈现给了每一个回答人(一般为40~60),以提高求得的估计值的精度。打分工作反映在结果变量上,结果变量是基本的数量、概率、一组无序或有序的类别。

## 析因调查法中的数据分析

分析方案的拟订始于对打分模式的审视,这样的审视可以在某些有实质内容的上下文中得到相当多的信息(例如据判断报酬过低和过高的工人的比例);然后继续进行信仰/判断方程的估计。总起来共有以下三种主要的方法:(a)协方差方法的分类分析法;(b)表面上似乎不相关的回归方法,但在这种方法中,来自特定回答人的方程的误差是相关的;(c)随机系数方法,在这种方法中,回答人构成了一个随机样本,且特定回答人的方程的参数视为是来自某种分布的。在所有这三种方法中,都有一个重要的步骤涉及回答人在内的同质性检验。

依据有实质内容的上下文和数据的特性,下一个步骤是估计决定性因素方程和结果方程。决定性因素方程可以和信仰/判断方程联合起来估计。多年以来,ROSSI法的实施多少有一定的困难,因为给出的计算资源需要生成情景模拟的总体,并估计特定回答人的方程。然而,近年来,随着桌面计算能力的提高,使得情景模拟的生成、随机样本的抽取和数据分析变得简单,因此,为新一代的研究使用ROSSI法打开了方便之门。

## 延伸阅读

上面摘录的罗西的文章,加上近年来进行的析因调查研究,给我们提供了有关数据收集和数据分析的程序全面深入的阅读资料。

——Guillermina Jasso

(沈崇麟译校)

## 参考文献

Jasso, G., & Rossi, P. H. (1977). Distributive justice and earned income. *American Sociological Review*, *42*, 639-651.

Rossi, P. H. (1951). *The application of latent structure analysis to the study of social stratification*. Unpublished doctoral dissertation, Columbia University.

Rossi, P. H. (1979). Vignette analysis: Uncovering the normative structure of complex judgments. In R. K. Merton, J. S. Coleman, & P. H. Rossi (Eds.), *Qualitative and quantitative social research: Papers in honor of Paul F. Lazarsfeld* (pp. 176-186). New York: Free Press.

Rossi, P. H., & Anderson, A. B. (1982). The factorial survey approach: An introduction. In P. H. Rossi & S. L. Nock (Eds.), *Measuring social judgments: The factorial survey approach* (pp. 15-67). Beverly Hills, CA: Sage.

Rossi, P. H., & Berk, R. A. (1985). Varieties of normative consensus. *American Sociological Review*, *50*, 333-347.

Rossi, P. H., Sampson, W. A., Bose, C. E., Jasso, G., & Passel, J. (1974). Measuring household social standing. *Social Science Research*, *3*, 169-190.

## 合成谬误(Fallacy of Composition)

合成谬误之所以是谬误在于没有认识到,对于整体的部分是真的东西,对于全体来讲可能未必是真的。这种谬误的一个特别的例子称为公地悲剧(tragedy of the commons),在这个悲剧中,土地是公共拥有的,集体上属于每一个人,而非任何一个特定的个人。因此,没有一个人关心这个财产,土地便因此而受害。一个日常的例子是一个推理:如果一个人站在自己的座位上能更好地观看芭蕾舞,于是就因此而认为,如果每个人都站在他(她)自己的座位上便都能更好地观看芭蕾舞。

——Tim Futing Liao

(沈崇麟译校)

## 客观主义谬误(Fallacy of Objectivism)

这种谬误的问题在于错误地假定实体独立地存在于我们的意识。在现象学【Phenomenology】中,分析者力图摆脱这样的假定。在社会科学中,它有可能取一种错误的假定形式,将我们概念上的范畴作为与理解社会现实的方法截然相反的存在。

——Alan Bryman

(沈崇麟译校)

## 可误论(Fallibilism)

可误论是研究者为反驳内在于其正在进行的研究中(方法论上和实质上)的可误性而进行的深思熟虑和坚持不懈的尝试。经过了可能的最严格的审查之后,如果有些方法论或结论还是相对完好无损,那么研究者就对研究的这些方面的信度【Reliability】和效度【Validity】就比较有信心了。因此,可误论把研究看成是研究者承担的一种不断进行批判性详查的义务。

作为一种研究方法,可误论吸取了三种

知识传统。它从科学哲学中拓展了可证伪性【Falsifiability】的原则。正如卡尔·波普尔(Popper,1965)明确指出的,可证伪性要求研究者把调查的形式和具体的假设置于最严格的审查之下,谨防其核心论点被推翻。从这个观点来看,做研究就与达尔文寻找适者生存的知识一般无异。其重点是失验(disconfirmation)、寻找误差、遗漏和矛盾。

可误论从实用主义【Pragmatism】那里借来了一个理念,即调查不是要寻找最后的普适真理,而是寻找视条件而定的、面向未来的活动,在该活动中,研究者承认,在任何一个瞬间获得的最好的知识,都存在着对其作进一步重新论述的可能性。实用主义有一种实验的、反基础性的味道,提倡不拘一格使用各种方法和路数来增进民主。它鼓励对假设进行持续的批判分析,对实践或调查的具体化、标准化的模式表示怀疑。实用主义特有的可误论的自我批判性(self-criticality)这一要素,意味着它从不盲从某种特定的方法论。

可误论从批判理论【Critical Theory】那里吸取了该传统对主导意识形态的怀疑论。批判理论感兴趣的是揭发意识形态——那个看似正常、自然、明显,并向人们解释社会如何运转的观念系统——并帮助维护不具代表性的少数群体的权力。它主要的知识工程——意识形态批评——致力于揭露人们是如何学会拥护那些最终会伤及自身的观念和行动的,而人们对正在发生的一切都不知情。如同可误论一样,批判理论也认为人们看作普适的规范、无可争辩的信仰、想当然的解释实际上都是人们构建的,都具有无限的可塑性。尽管可误论不像批判理论那样重视揭示主导意识形态是如何把不公正的社会秩序合理化的,但在对传统真理保持怀疑这一点上,它们是一致的。

可误论的定位就是要去揭示研究方法或发现中有什么是不确定的、未经检查的和自相矛盾的。这类研究的范例是加里·凯尔(Cale,2001)在社区大学的大学生中所做的研究,在该研究中,他分析了鼓励大学生进行批判的、反向的思维的尝试实际上是如何走向它的反面的。

——Stephen D. Brookfield

(杨可译　沈崇麟校)

### 参考文献


Bronner, S. E., & Kellner, D. M. (1989). *Critical theory and society: A reader*. New York: Routledge.

Cale, G. (2001). *When resistance becomes reproduction: A critical action research study*. Proceedings of the 42nd Adult Education Research Conference. East Lansing: Michigan State University.

Cherryholmes, C. H. (1999). *Reading pragmatism*. New York: Teachers College Press.

Popper, K. E. (1965). *The logic of scientific discovery*. New York: Harper & Row.


## 证伪主义(Falsificationism)

证伪主义是一种科学哲学,也被称为批判理性主义,它用演绎逻辑来为假说演绎法【Hypothetico-Deductive Method】提供基础。证伪主义 1930 年代由卡尔·波普尔(Popper,1959)提供基础。它是 1930 年代波普尔开发出来处理实证主义的缺陷的。它拒绝实证主义立场,以便于使用一种基于批判的试错法的不同的解释逻辑。在这种

方法中，理论都要对照“真实”进行测试。证伪主义与实证主义的本体论【Ontology】在某些方面有共通之处，但它拒绝实证主义的认识论【Epistemology】。它的立场是，自然科学和社会科学的差别只是它们的内容，而非它们的方法的逻辑形式。

尽管波普尔不属于维也纳学派这个发展出了逻辑实证主义的科学家和哲学家群体，但他和这个学派在知识上有密切的接触。他也接受维也纳学派的传统，认为科学知识可能不完美，但却是人类能获得的最确定和最可靠的知识。不过，他对实证主义，尤其是逻辑实证主义却很不满，想方设法使自己与这一学派保持距离。他反对所谓观察给科学理论提供了基础的观念，并承认形而上学观念在科学理论形成过程中起着重要作用。

波普尔认为，观察是在演绎推理时使用的，理论发明出来就是要解释观察，而不是像在归纳中那样来源于观察。科学家们不应该等着自然来揭示其规律，而应该把规律应用于这个世界，通过试错的过程，用观察来尝试驳斥错误的理论（演绎）。经过这一关键的过程仍能幸存的理论就暂时被我们接受，但永远也不能被证明它就是真理。所有的知识都是试验性的，并要接受不断的批判性评估。这种批判态度采用了口头辩论和观察；观察正是为了辩论而使用的。

究竟先有理论，还是先有观察，在波普尔看来是不成问题的。他承认一个假设是以观察为先导的，尤其是那种追求解释的假设。不过，这些观察要预先假定一个参照框架，也就是一个或多个理论。

波普尔在后来的著作中，在寻觅社会科学适用的逻辑时，总结了他的主要论点。社会科学的逻辑，就像自然科学的逻辑一样，存在于对某些问题试着找出试验性的解决方案中，解决的方案被提了出来，并接受批评。一个不能批评的解决方案必须被排除，因为它是不科学的。终极的批评就是试着驳倒理论。如果理论被驳倒了，必须尝试另一个理论。不过，如果理论经得起检验，我们可以暂时接受它，尽管它还必须经受进一步的讨论和批评。因此，科学的方法是一种被严格的批评控制的猜想，是用来解决问题的试验性尝试。所谓的科学的客观性就在于批评方法的客观性。

波普尔认为，尽管科学是寻找有关世界和宇宙的真理的，但我们却永远也不能确定这些理论是否正确。我们能做的就是通过猜测和反驳的过程来排除错误理论。有些理论会被抛弃，有些会被暂时接受（被证实）。这个过程让我们得以尽可能地接近真相，但我们永远不知道在什么时候我们已经建立了正确的理论。未来的某位科学家也可能会在其他环境中检验这个理论，并发现它是错误的。因此，理论必须一直被视为暂时性的——它们可能在将来得到完善或被驳倒。我们能做的就是除掉那些不符合实际的理论。波普尔认为，没有比推测和反驳更理性的步骤了。这是一个这样的过程：大胆地提出理论，尽我们最大的努力来展现其错误，如果我们的批判未果，那么我们就暂时接受它。

用演绎逻辑构成的批评也适用于证伪主义。

——Norman Blaikie

（杨可译　沈崇麟校）

## 参考文献

Blaikie, N. (1993). *Approaches to social enquiry*. Cambridge, UK: Polity.

Chalmers, A. F. (1982). *What is this thing called science?* St. Lucia: University of Queensland Press.
Popper, K. R. (1959). *The logic of scientific discovery*. London:Hutchinson.
Popper, K. R. (1961). *The poverty of historicism*. London:Routledge & Kegan Paul.

## 女权主义认识论(Feminist Epistemology)

女权主义认识论是由贯注了女权主义的认识论原则组成的。

——Alan Bryman

(杨可译　沈崇麟校)

*也可参见女权主义研究【Feminist Research】、立场认识论【Standpoint Epistemology】。

## 女权主义民族志(Feminist Ethnography)

女权主义民族志【Ethnography】是贯穿了女权主义研究【Feminist Research】原则的民族志。因此,女权主义民族志记录的是女性的生活和活动,从女性的视角进行研究和分析,并重视情境中的女性立场和观点。由于很难真的按照女权主义原则来办,人们对女权主义民族志的前景仍有争议,因为难以以完全非利用的方式(nonexploitative way)来行事。针对这种观点,女权主义研究者提出,应该在研究者与参与者的互动和交往中注入互惠原则,这样就能弥补研究者及其所研究的女性之间的利用关系。

——Alan Bryman

(杨可译　沈崇麟校)

## 女权主义研究(Feminist Research)

女权主义研究是作为对西方社会科学中已经察觉到的两个相关缺点的一种反应而发展起来的。第一个缺点是相对地无视女性,缺乏对影响女性生活的议题的关注。第二个缺点涉及建构知识的社会研究的实践及其过程。它认为,此前对社会世界的研究一直都是从男性利益和关怀的视角出发,在研究女性生活和看待世界的方式时无视其中出现的不同。被当作中立、客观、价值无涉的知识原来只不过是不公平的、性别化的知识。女权主义研究的目标就是要把女性的经验更充分地带入人们的视野之中。这要求我们挑战传统的研究习惯,同时要从根本上重新评价许多有关社会科学本质的想当然的假设。其总体目标就是要更好地理解性别不平等的本质。

然而,并不是关注女性、性别或性别化的生活本身就让一个研究项目成为女权主义的项目。首先,女权主义的研究之所以被定义为女权主义的,是因其与女性相关、由女性从事和为女性服务的研究。不过,近来有人建议,识别女权主义的研究方法可以看其有关性别和权力的理论架构、有关公正的框架和概念(无论是如何构想的)、对转型和社会变迁的关注以及有关伦理和责任性的观念(Ramazanoglu, 2002)。有关这些议题的观点并不统一,而识别一个研究是否是女权主义研究也从来就不是一目了然的事。

女权主义方法和其他社会调查方法之间，也有部分定义性特征是相互重叠的。

女权主义者早前在有关社会研究的讨论中，就对方法、方法论和认识论【Epistemology】进行了区分(Harding, 1987)。方法是用来指称进行研究的技术和工具。方法论所指的是这些方法实质上是如何操作化并付诸实践的。认识论关注哲学讨论，确定哪些知识是可能的，如何保障这些知识是充分且合理的。女权主义者一直都在精细化其对于这些议题的看法，而关系这些议题的广泛的争论也一直没有中断过。

## 研究的方法

1980年代早期，大量关于女权主义的争论都围绕着什么构成了为我们所用的合适的方法这一问题。人们对当时用于调查研究以及强调科学的测量【Measurement】和客观性【Objectivity】的种种方法是否更具合理性提出了疑问。有人认为，这类研究割裂了人们的生活，造成了原子化事实的生产和测量，而这些事实的重要性在该研究开始之前就确定了。针对女性展开的研究，尤其是针对她们那些相对隐秘的领域展开的研究就被认为更有问题。例如，诸如暴力、性、母亲身份等议题，我们已有的知识很少，无法根据这些知识写出要调查的问题。

因此，许多女权主义者提出，基于阐释论【Interpretivist】哲学立场的定性方法以及使用敏感、灵活、开放的方法来生成数据比定量方法更为合适。诸如深度访谈【In-Depth Interview】、生活史访谈【Life History Interview】和民族志【Ethnography】等方法关注被研究事物的意义和阐释，因此，使研究者得以通过参与者的眼睛来看这个社会世界。也有人感觉定性方法更契合女权主义的原则，可以更促进平等，反对物化和主从关系。这些主张在女权主义研究者中成了某种正统，尽管即便在当时，也并不是每个人都赞成这种立场。比如说，意识到性暴力或者单亲母亲发生贫困的普遍性，在一定程度上取决于是否能获得说明问题严重性的数据。

近年来，有关定性研究【Qualitative Research】的古老正统已经瓦解。有些评论家称，这些方法可能比定量方法更具有剥削性，因为它们有赖于信任和融洽关系【Rapport】的假定，这可能会掩盖权力和不平等。女权主义者采取了一种更为务实的方法，她们倾向于融合各种方法，甚至还采用了定量方法，以便开发可以用于政策发展的更为解放的社会科学。事实上，有些人建议，可能在有些情况下(如调查性侵犯)，被调查者可能觉得完成调查问卷比接受访谈压力更小。不过，经费和资源的限制常常意味着相对小规模的定性调查项目，仍然是大多数女权主义者得以开展研究的主要方法。

## 方法论

有关研究方法的辩论关涉的第二个议题与女权主义方法论或研究实践有关。有人建议，女权主义研究和其他形式研究的区分主要不在于方法本身，而在于它们所问的各种各样的问题，研究者和研究及其实施的关系，以及工作背后的意图和目标。其中特别关注的问题是如何保障研究以非剥削的和有道德担当的方式来开展。其中心议题一直都是研究者与参与者之间是否可能存在结构性的不平等，以及这种不平等可以缩小到何种程度。有人始终认为，女权主义者在他们的研究中也复制了他们在别处所批评的权力关系，因而他们是虚伪的，这种行为损害了生产出来的知识。

处理这些议题的一系列策略一直都有争论，其中有些策略渗入了更为主流的非女性主义研究实践。举例来说，这些策略包括

在研究中与参与者发展融洽关系、充分交流,尽可能地告知他们研究的内容和目标;尽可能诚实、敏感地回应参与者的疑问、问题和担忧;提供可以提供任何必要信息或社会支持的组织或出版物的名字给他们。有些女权主义者主张,在阐释和分析研究数据时,要得到被研究者的反馈,以确保所听到的妇女的声音没有被歪曲和剥削。重点在于在研究项目的设计和实施中,应将参与者也包括进来,同时也要保证研究发现在相关人等中广泛传播。

不过,尽管有女权主义研究者似乎认为在研究过程中能去除大部分权力关系,但许多人并不那么乐观。在放松的环境中,受过教育的研究者说服参与者开口大聊他们的生活太容易了。但参与者可能会认为,当初还是以保持沉默为好。研究者很容易就认为他们已经建立了和谐、平等的关系,而他们所研究的人们的看法不尽如此。有些女权主义者也提出,研究者可能被赋予了过高的权力,因为被研究者都被归于一个被动的、被牺牲的、没有明显的反抗空间的位置。

对研究中的权力关系的议题的一种处理方法,是借助反身性【Reflexivity】这一概念,我们甚至可以说,这已成了女权主义方法论所关注的核心原则。反身性有两个不同的方面:第一个方面是要讲清楚,尽管已尽力减少权力的影响,权力可能还会怎样发挥作用、影响研究过程。其关注重点在于构建研究的假设和伦理判断,正是它们设定了研究的框架,以及研究计划和过程的构建,从而展示研究者为何能解释其生产的知识(Ramazanoglu, 2002)。其理念就在于把社会研究的批判方法用于社会研究实践本身。

第二个方面是反身性探索研究者的社会位置。这指的是研究者的个人生平和位置与她所参与的项目有相关性。其论据是研究者无法与其工作分离开,因为研究工作离开了研究者的参与就无法存在。因此,在任何讨论、分析或对研究结果的阐释中,包括对研究者和研究的关系的批判性反思是非常重要的。

实际上,女权主义者称,这种对研究的双重反身评论,对知识的阐释是不可或缺的。他们批判所谓的"弱的"客观性,致力于净化带着各种价值观的知识,以及建立在这些知识之上的研究(Harding, 1991)。这种客观性的形式是虚弱的,因为这些价值观假装不存在,实际却以隐匿和未予明言的方式包括在内了。相比之下,所谓"强的"客观性,则必须对所有作为研究过程列出的证据都进行批判式的考察(Harding, 1991)。强客观性及其对反身性的强调包括了对背景观念和文化动机的系统考察。在研究过程中设置这一部分本身就说明,生产的知识变得更严格、稳健和可辩护的了。

不过,对于反身性和强客观性实际上可能实现多少还是有一些争论的。在过程中保持反思意味着时间拉长,而且也很难确定关于这个过程持续多久、走多远的指导纲领。还有一种危险就是,对研究过程及研究者的涉入的叙述将会主导研究分析和发现。这可能会导致参与者生活的复杂多样简化成了研究者本人的自传。

关于差别的概念对女权主义的方法论观点有着极大的影响。早期的假定有一种基本的女性气质,认为女性的压迫是普遍的、同质的、共享的。这些假定一直在受到挑战。对差异的强调把注意力转向了女性之间的多样性,尤其源于民族、性、残疾、阶级和年龄的多样性。这意味着,仅仅是作为女性来研究女性并不足以建立融洽关系和削弱社会等级。这些其他的经验维度也可能会影响研究过程,这就是为什么要建议可以把研究者和被研究者在共有特征上进行匹配。不过,要明确哪些差异可能会对某个特定研究产生影响、这种影响的性质严格说来到底是什么,并不总是那么简单。那么,

总体而言,共有的性别特征自身并不能推进共识,不能假定仅依靠性别本身就能轻而易举地建立起信任来。

## 认识论

女权主义对认识论的关注核心就在于谁知道什么、关于什么人,以及这种知识怎样才算是合理的。早期的作家对那些只在已有的框架中研究妇女的人多有批评。为了让女性生活的意义更明显,有必要从女性的角度来进行分析。这就形成了所谓的立场认识论【Standpoint Epistemology】。对于采取女性主义的立场有许多不同的思考维度,但核心就是女性特殊的性别经验会产生不同的和特殊的理解。这种通向不同知识的通道,使揭示男性立场可能看不到的人类关系的存在形式成为可能。因此,立场认识论给我们提供了一种新的且更为可靠的洞察女性生活的可能性,因为它就是以女性经验为基础的,包括情感和体现。

立场认识论还以知识唯实论【Realist】方法为基础;认为系统或结构因素影响生活经验,而这些结构是可知的。女权主义研究的重点就是不断地生产对女性生活的更好的理解。不过,随着后现代思想的发展,这种观点已遭到了挑战,立场认识论常常被拿来和后现代叙述相比较。

与后现代主义【Postmodernism】一样,女权主义思想也有不同的形式和重点。不过,它和大多数(非女权主义的)变化形式有许多共同特征。它们对于一般化、普适的范畴、稳定连贯的自我的存在、语言的清晰性以及科学理性生产"真理"的能力都有批评(Nicholson, 1990)。其焦点是话语、文本性、碎片、多重主体性以及不确定性。立场认识论承认,各种叙述都是选择性的建构,同时也认同这些叙述可能在一定的严谨程度和信度之下代表独立的现象;而女权主义后现代主义则否定这一点。有人宣称,女权主义后现代主义已经把所有女权主义方法论的基础都一扫而空了(Ramazanoglu, 2002)。正是因为社会现象只能通过话语的使用来理解,因为话语实践本身就是要激发各种社会性的形式,因此就不可能不在理解社会世界的同时也构建社会世界。于是,后者就成了一个它意味着"知道"什么的任务。重点就从试图在对女性经验的叙述中考察社会关系的本质转向了话语和文本,而各种事物正是借助话语和文本才被知道的。实际上,经验这个概念本身,正如人们所拥有的某个东西一样,是否可以根据表面来判断也成了有问题的了。所有这些都会给批判地考察社会世界本质的社会研究者带来困难;从后现代的角度来看,这是不可能的。尽管没有否认他们自身的物质表现,也即他们生活在物质世界中,但女权主义后现代主义者宣称经验和物质的现实之间的特殊连接是无法详细说明的(Ramazanoglu, 2002)。而女权主义的研究就转变成了不断地解构认同、主体性、自我,接下来这些解构又要被进一步解构。

后现代思想对女权主义研究有着巨大的影响,有人甚至认为任何开展经验工作的企图都是误导和有缺陷的。其他人则认为,采用来自后现代方法的部分洞察还是可能的,尤其是在研究文化、意义及象征的模式时。对女权主义社会研究者来说,问题就是如何在具体的和物质的差别、权力和不平等经验性的关注点,与对如何生产知识的批判性反思之间保持平衡。而是否能把社会调查的经验方法和对主体性和文本/阐释生产之间的关系的关注结合起来,则仍是一个充满争议的问题。这个问题在女权主义者中间产生了差别,而她们对这个问题的回答则各有千秋。

——Mary Maynard

(杨可译　沈崇麟校)

## 参考文献

Harding, S. (Ed.). (1987). *Feminism and methodology*. Milton Keynes, UK: Open University Press.

Harding, S. (1991). *Whose science? Whose knowledge?* Milton Keynes, UK: Open University Press.

Hartsock, N. C. M. (1998). *The feminist standpoint revisited and other essays*. Boulder, CO: Westview.

Nicholson, L. (Ed.). (1990). *Feminism/postmodernism*. London: Routledge.

Ramazanoglu, C., with Holland, J. (2002). *Feminist methodology*. London: Sage.

# 虚构与研究(Fiction and Research)

在20世纪中期以前,社会研究在确定虚构的文学作品的社会背景时,倾向于减少虚构——这是进化论、结构功能主义和马克思主义中实证主义【Positivism】的遗产。与其不同的另外一种研究方法是文学研究方法(literaturwissenschaft approach),这种方法通过分析虚构的作品来了解它产生和指称的社会,并用文化研究来对虚构的作品进行结果分析(Milner, 1996)。它结合了两种不同的研究方法:外在方法将虚构作品看成它生产、分配和接受的社会文化条件的批判性的反思来分析;内在方法则将它作为一种它对社会的反身性的建构特征来分析,其中文本【Text】的语言结构,可以作为社会关系的结构同源性来加以分析。虽然内在方法强调文本语言的精致和复杂,力图避免在对社会现实进行例解时减少虚构,但是它们将在复杂的现代社会中多数人经验的代表性这一研究问题上提升了高度,并将其更有效地应用于少数人的、有选择性的文化传统和先锋现代主义。外在方法对通俗小说提供了更有效的分析,因为此类作品很少持久地关注与其他文化背景和历史时期的社会的经验对比(Loewenthal, 1989)。

现在,更多将语言和社会结构的关系作为文本的建构因素的社会语言学和新马克思结构主义【Structuralism】研究(Filmer, 1998),则关注虚构文学作品的语言是如何指明可能的阅读的诠释学【Hermeneutic】问题。这需要对文本既能创造自身又能指称现实的能力保持开放态度。以下的研究方法反映了这一点。

## 符号学和结构语言学

符号学【Semiotics】和结构语言学【Structural Linguistics】提供了关于符号如何使文本中的语言产生意义,以及它与日常生活语言和其他批判性话语之间关系的理论(Eco, 1981)。

## 现象学社会学

现象学社会学分析了惯常表达和文本中体现的社会现实,以及读者的体验和话语世界的外部现实这几方面意涵之间的关系(Schutz, 1964)。

## 情绪结构

情绪结构分析文本中潜在的符号学和符号学结构,并将其看成那些随着变化的情境而自然产生的新体验、预示着新的社会结构的表达(Prendergast, 1995)。

## 遗传学结构主义

遗传学结构主义将结构学派的马克思主义分析应用于虚构文本,明确有力地表达

自然形成的社会群体对世界的认知。他们分析了这个共享意识在促使群体不断拓展新的共享意识,进而改变现存的社会制度和政治秩序方面所具有的作用(Goldmann,1981)。

所有上述研究方法都用虚构文本作为一种语言、主观想象和社会结构之间的映射关系,以精确地将虚构文本建成社会研究的重要议题资源,这些议题包括个体和社会之间关系的结构、社会变迁过程、社会行动的意义和原因。通过这样的做法,这些方法试图超越个人传记、特定文化和历史时期的特异性。

——Paul Filmer

(郑晓娟译 沈崇麟校)

## 参考文献

Eco, U. (1981). *The role of the reader: Explorations in the semiotics of texts*. London: Hutchinson.

Filmer, P. (1998). Analysing literary texts. In C. Seale (Ed.), *Researching society and culture*. London: Sage.

Goldmann, L. (1981). *Method in the sociology of literature*(W. Boelhower, Trans. & Ed.). Oxford, UK: Basil Blackwell.

Loewenthal, L. (1989). Sociology of literature in retrospect. In P. Desan et al. (Eds.), *Literature and social practice*. Chicago and London: University of Chicago Press.

Milner, A. (1996). *Literature, culture and society*. London: UCL Press.

Prendergast, C. (1995). *Cultural materialism: On Raymond Williams*. Minneapolis: University of Minnesota Press.

Schutz, A. (1964). Don Quixote and the problem of reality. In *Collected papers*, Vol. 2. The Hague: Martinus Nijhoff.

Williams, R. (1977). *Marxism and literature*. Oxford, UK: Oxford University Press.

# 田野实验法(Field Experimentation)

田野实验不同于实验室实验【Laboratory Experiment】,是发生在自然场景下的随机干预。这些干预措施的性质存在巨大差异。这样的例子包括随机的教育券项目、选民动员运动、警察对发出爆裂声的房屋的突然搜捕、就业培训项目、所得税时间表和健康保险计划。田野实验法的分析单位也各不相同。大部分田野干预以个人为目标,但也有许多研究是对群体或组织进行随机干预。

与民族志研究或描述性研究不同的是,田野实验法主要致力于建立因果【Causal】关系。一个执行良好的田野实验结合了随机设计和外在效度【External Validity】两个方面的优势。例如,要研究商业广告的影响,实验室研究通过在人造场景下对不同的控制组【Control Groups】施加不同的广告刺激,然后在模拟市场的情境中测量他们的购买偏好;田野实验则是对真实的广告活动中的内容和时机进行随机控制,并试图将这些不同的干预措施与观察到的消费者需求类型联系起来。两种类型的研究都使用了随机化,但后者的优势在于能够在直接的、具有现实世界适用性的情况下建立因果关系。

田野实验法在项目评估中也是一个非常有价值的工具,特别是用于检测新项目或是现有项目的不同版本的效果。例如,假设研究者想要评估一项以学校为单位的、药物预防项目的有效性。有些学校可能接受了药物(滥用)提醒项目,而有些学校却可能没有。学校接受药物提醒项目的顺序会有随机变化,或者整个学校总体可能接受了另外的药物提醒项目。通过比较实验组和控

制组的结果,评估者可以对干预的效果作出无偏的评价。

自然场景下的实验具有一个十分有吸引力的特征是数据分析的透明性。非实验的数据分析结果在很大程度上取决于研究者对数据所使用的模型,因而研究者通常会拟合许多模型,努力找到一个合适的模型,与此不同的是,实验的数据分析是非常稳健的。控制组和实验组的简单对比通常足以给出对实验组结果的无偏归因,对协变量【Covariates】的额外分析只不过是在更大的精确度上估计因果参数。这并不是说实验研究不需要数据挖掘,但是由原始数据向统计数据的转化包含了更少的风险,因而面临的道德困境也更少。

从原则上讲,田野实验法代表了社会科学家能够获得的可靠的因果推论的最强根据。但实际上,这个方法论面临诸多重要的伦理和操作限制。社会科学家很少有机会控制他们最感兴趣的变量,例如文化、政治系统、经济繁荣等。即使在随机干预能够实现的情况下,有时也会因为执行这些措施的人,而导致干预的失效。怀有良好意图的内科医生将病情最严重的病人偷偷带进实验组的行为,可能会导致实验产生有误导性的结果。

精心策划实验研究和维持随机过程的完整性所面临的挑战意味着,实验最好发生在为数不多的几个地点,而这些地点的选择更多是出于后勤保障的考虑,而非代表性的需要。这个限制带来了另一个问题,即研究结论是否只适用于那些实际参加实验的人员类型。一般来说,对实验的复制【Replication】可以有效地回应那些针对特定时间、地点和人群的实验所得出结论的疑虑。

如果实验的参与对象拒绝参与研究或者无法接受实际的实验干预,这时又会产生进一步的难题。尽管不服从和实验损耗会降低实验设计【Experimental Design】的效力,但只要参与的决定与实验效果的强度无关,它们仍然是可补救的。统计修正采用一个工具性的变量回归,在这个回归中,自变量是参与者是否真的接受了实验的实际干预,工具变量【Instrumental Variable】是参与者是否本来就被分入了处理【Treatment】组。产生的效果估计值被用来表示"实验对被实验者的效果"。实验干预的效果是否真的有区别,能够通过实验性的操控参与率来进行研究(例如,通过改变给参与者提供的鼓励)。

由于这些操作性和伦理因素的考虑,社会科学家经常转向自然实验【Natural Experiment】,在该方法中实验组和控制组以类似于随机分配的方式形成。例如,如果某市领导人在选举年扩大警力规模,以安抚对犯罪较为关注的市民,这样选举周期会提供一种用来研究警力对犯罪率影响的工具。类似地,如果公立学校以学生的出生日期为依据来接收入幼儿园的学生,我们就可以通过对比在入学门槛日期前后的孩子,来研究幼儿园录取对智力发育的影响。与田野实验法类似,这些研究方法抓住机会去研究自然场景下,一些近似的外生变量的后果。但与实验法不同的是,自变量的外生性无法通过随机过程程序予以确保。

——Donald P. Green

(郑晓娟译　沈崇麟校)

## 参考文献

Angrist, J. D., Imbens, G. W., & Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, *91*, 444-455.

Cook, T. D., & Payne, M. R. (2002). Objecting to the objections to using random assignment in educational research. In F. Mosteller & R. Boruch

(Eds.), *Evidence matters: Randomized trials in education research*. Washington, DC: Brookings Institution.

Green, D. P., & Gerber, A. S. (2002). Reclaiming the experimental tradition in political science. In I. Katznelson & H. V. Milner (Eds.), *Political science: State of the discipline* (3rd ed., pp. 805-832). New York: W. W. Norton.

Heckman, J. J., & Smith, J. A. (1995). Assessing the case for social experiments. *Journal of Economic Perspectives*, *9*, 85-110.

Rosenzweig, M. R., & Wolpin, K. I. (2000). Natural "natural experiments" in economics. *Journal of Economic Literature*, *38*, 827-874.

## 田野关系(Field Relations)

社会科学田野研究的明确特征是研究者和被研究者之间的相遇。田野中的相遇有不同的方式,有些关系持续的时间较短,有些则持续了足够长的时间,构成不断推进的关系的基础。在后面这种情况下,研究者和被研究者之间的关系性质对研究进展和成功与否起着核心作用。

或许最明显的观点是,所有集中的田野研究都需要与各种不同的人细心和设身处地维系持续良好关系。田野关系中两个最重要的关系是与看门人和关键知情人的关系。如果长期的田野研究想要成功,这两个关系是非常重要的,或者应该是不可避免的。看门人能够促成对田野场景最初的调查通道【Access】和对处于其中的信息、情境和个人的初次接触,而他们的持续支持在整个研究阶段也仍然具有重要性。典型的看门人包括政府官员、档案保管员、组织中的高级管理人员、社区领袖和非正式网络中的核心成员。关键知情人,其原型来自威廉·怀特(Whyte,1955)写的《街角社会》(*Street Corner Society*)一书中的"Doc",他可能也是看门人。关键知情人常常都是看门人,但反之则未必。如果研究者缺乏使用当地方言进行调查工作的能力,这时关键知情人和看门人的作用就会变得极为重要。

与看门人和关键知情人重要性相关的问题是如何找出最有用的看门人;如何防止自身对看门人和关键知情人的过分倚重;如何防止看门人对研究的专擅和控制;如何避免过分听信于某些人,一般都是利益相关的关键知情人提供的情况,而忽视其他研究参与者的观点。所有这些问题都不存在简单明了的、教科书式的解决方法。

当研究者本身已经有了正在调查的社会情境的成员资格角色时,田野关系会显得格外重要(Adler & Adler,1987)。这样的资格身份能够为调查带来很多益处,包括更容易进入调查,以内部身份能够获得更多全局或局部的信息。今天的社会科学调查,特别是评估型的工作,这种现象正越来越普遍。但是这种情况下也有许多值得注意的问题,包括保持评判距离和客观性,处理其他成员对个人和组织利益的需求,平衡伦理责任,以及考虑以后如何以非研究者的身份再次进入等。

最后需要指出的是,如果在最为普遍和日常的意义上,田野研究工作者和被研究者之间的关系对研究结果是至关重要的话,那么在日常的人类世界中,性别关系便是十分重要的(Warren & Hackney,2000)。考虑到日常生活,无论任何场景或文化都是围绕性别关系组织起来的,上述结论并不令人吃惊。无论研究主题是否与性别相关,在开展

田野研究时什么人能够去哪里、和谁一起、谁可以与谁谈话,这些都是对开展研究和获得研究结果具有重要影响的性别关系问题。

——Richard Jenkins

(郑晓娟译　沈崇麟校)

## 参考文献

Adler, P., & Adler, P. A. (1987). *Membership roles in field research* (Qualitative Research Methods Series vol. 6). Newbury Park, CA: Sage.

Warren, C. A. B., & Hackney, J. K. (2000). *Gender issues in ethnography* (Qualitative Research Methods Series vol. 9, 2nd ed.). Thousand Oaks, CA: Sage.

Whyte, W. F. (1955). *Street corner society* (2nd ed.). Chicago: University of Chicago Press.

# 田野研究(Field Research)

"田野研究"或"田野工作"这种表述,最初来源于自然科学,指的是在人为控制的实验室环境以外的自然场景中收集研究资料。在我们的语境中,尽管近年来我们过分强调了田野研究与民族志研究和其他定性方法的关系(例如,Burgess,1982;Whyte,1984),但其实大多数的社会科学研究可能都与这个词有关。无论是作为对大规模定量社会调查【Survey】的替代,还是通过参与观察【Participant Observation】达到长期亲密关系,社会研究在涉及基本资料的收集时都会引入一些不同类型的田野工作。我们可以将田野工作界定为研究者和被研究者在非控制环境下的相遇(档案研究【Archival Research】、传媒研究和焦点小组【Focus Group】研究可能是明显的例外)。

为什么社会研究者应该如此倚重田野工作?或者更直接地问,为什么社会研究者很少使用实验方法?对这个问题没有单一的答案。主要的社会学传统的伦理可能对人体实验给出了一个关键限制(尽管一些在伦理方面十分敏感的学科,如医学和心理学仍然依靠实验证据,这个事实表明伦理问题不可能是决定性的)。或许,另外三个因素对这个问题更有解释力。

第一个因素是难以在真实场景之外,创造足够可信的社会研究环境(除了一些特定的小规模和具有明确限定边界的场景)。模拟毕竟只是模拟,它可能与我们对人类世界日常生活的理解关联很少。第二个因素也与此相关,在社会研究情境下,难以维持被适当控制的实验环境。这些问题都意味着对大多数社会研究的目的而言,不存在可以替代自然论【Naturalism】的社会研究。

第三个因素可能也是最重要的因素,是可预测性问题。人类是具有反射性的生物,那些为社会研究所关心的人类行为的方方面面,都受到情绪、理性决策、习惯和冲动等因素的综合影响。因此,人过去一直在做的行为无法成为预测未来行为的依据。由于实验方法的目标是对某一事件和事件发生的控制条件积累足够精确的观察数据,进而作出可靠的预测(例如,公式化的表述和概率论的推论),因此它在社会研究中的应用肯定会受到一定的限制。

然而,这并不意味着在田野研究中不存在类似实验的方法。特别是结构化访谈【Structured Interview】与实验方法具有一些共同点,采用这种方法的目标是发现在面临以同样方式或尽可能相近的方式提出的相同问题时,被研究者的样本中会作出怎样的

回答。严密的抽样方法也可以看作与实验控制方法的另一个近似之处。此外,像"如果……怎样……"式的投射性问题也带有明显的实验方法特征(当然它们与哲学家口中的"思想实验"也具有某些共同点)。

尽管如此,由于田野研究必然是非控制性环境的,以及社会研究所关注的人类行为具有难以把握的特征,它与实验室实验是具有很大差异的。从认识论的角度,田野研究也具有一些独特的特征。最明显的特征是对于研究资料的认知和记录,田野研究方法高度依赖研究者而不是研究工具。即使使用最为结构化的程序或步骤,研究者本身也才是研究工具。其结果是,使我们很难明确给出数据收集阶段结束和解释阶段开始的界限。

这还意味着在田野研究中,方法和数据的界限有时也是模糊的。例如,当哈沃德·纽比研究英国农场工人时,他能够找到和接触这个劳动群体样本的唯一方法,是通过他们的雇主。这一事实不仅是研究方法中的一个有趣细节,同时也成为显示他们社会隔离程度的指标(Newby, 1977)。与此类似,女性研究者通常可以更容易地接触年轻男性的生活,而男性研究者进入年轻女性的生活却有较大难度。这一现象也说明了在日常生活中,公共领域和私人领域性别的不对称特征。

这些例子进一步表明:田野的性质、范围和边界并非总是一成不变的。除非我们谈论的研究场景在范围和/或制度层面已经被清晰限定,例如,开展一些经典民族志研究中的小型住宅区和工作场所,否则确定研究要面对的田野,是田野研究工作中最重要,但通常也是最复杂的首要任务。除了已经提及的样本和性别问题外,规模和复杂程度也是需要考虑的因素。这些因素会决定采用哪种方法最为合适,以及采用每种方法的成本是什么。例如,现今的大多数场景都过于广阔、异质性强和多面向,使得研究者难以通过参与观察的方式"全面"进入(Jenkins, 1984)。一个考虑缜密的田野研究,通常需要其使用的研究方法,具有一定程度的灵活性和开放性。

研究者在田野中能够有怎样的发现还取决于研究过程的政治。考虑到这一因素,涉及的关键问题包括研究者身份、看门人角色和对特殊资源的依靠(或过度依靠)。研究者是谁,通过谁才得以进入研究场景,将谁作为关键知情人,研究者可以与谁交谈,哪些人会较为配合地回应研究者提出的要求,被研究者代表哪种地方性观念,所有这些因素都会影响研究得到的结论(见田野关系【Field Relations】)。

既然存在上述的许多问题,为什么社会研究者仍然继续使用田野研究法呢?让我们回到最初的观点,那就是除了几个明显例外,社会研究中主要的资料收集并不存在替代性方法。如果社会研究仍然致力于关注和研究人类世界及其变化,田野研究就仍然是必需的研究手段。档案研究、媒体研究和焦点小组方法的贡献都是有限的,无论我们采取哪种方法,都无法替代在人们的日常生活场景下与他们相遇这种方法的使用。

——Richard Jenkins

(郑晓娟译　沈崇麟校)

## 参考文献

Burgess, R. G. (Ed.). (1982). *Field research: A sourcebook and field manual.* London: Allen & Unwin.

Jenkins, R. (1984). Bringing it all back home: An anthropologist in Belfast. In C. Bell & H. Roberts (Eds.), *Social researching: Politics, problems, practice* (pp. 147-164). London: Routledge and Kegan Paul.

Newby, H. (1977). In the field: Reflections on a

study of Suffolk farm workers. In C. Bell & H. Newby (Eds.), *Doing sociological research* (pp. 108-129). London: Allen & Unwin.

Whyte, W. F. (1984). *Learning from the field: A guide from experience*. Beverly Hills, CA: Sage.

## 田野笔记(Fieldnotes)

田野笔记构成了民族志研究或参与观察研究中的民族志【Ethongraphy】记录,它被认为整个田野研究活动中最重要的部分。它是田野工作经历中按时间顺序形成的日志,包括对人物、事件、场景构成的描述;包括研究者与人们的对话、观察到的互动过程、事件的进展和持续时间;也包括研究者与调查活动相关的经历和体验。观察者记录所有与资料构成或资料质量相关的自身状态也是十分重要的。这些记录还可能包含对所观察事件和互动过程的解释以及对所研究文化的一些洞见。总之,田野笔记构成了分析由此形成的内容。

田野记录的文字应尽可能完整、全面,并尽可能快地形成。研究者通常认为,田野笔记越早记录,就越生动翔实。但是,在大多数情况下,马上记录观察结果是不可能、不方便和不合适的。在这种情况下,通常会先做些临时记录,为之后更详尽的笔记提供基本的素材。这种包含了短语、单词和不连贯句子的简缩笔记,可以快速甚至较为隐蔽地记录,这为日后重构该事件提供了能触动相关记忆的帮助。然后,研究者可以通过填充细节,以及回忆那些当时没有记录下来的情况来扩充这种简缩版笔记。需要明确的是,田野笔记的全面性只是相对意义上的。随着研究者对处理的数据的范围,以及提出的假设和解释变得更为熟悉,即使是已经完成的田野笔记也仍然可以进行调整、添加、修正和更新。

作为一项需要投入很大精力和相当程度的自律精神的活动,某些田野笔记最初可能看上去似乎过于表面化和琐碎无意义。但是,研究者不应当预先假定哪种信息要进入研究范围。至少在最初阶段,在研究者的意识中,应该认为所有发生在田野中的事情都可能是重要的,因此也都是值得记录的。随着研究变得越来越专注,研究者在什么是应当记录这一问题上也会变得越来越挑剔。

由于研究者在观察和记录什么和使用的特定词汇及术语方面具有选择性内容,因此其必须使用的语言,即使是无意而为,也会将一种结构施加于研究的社会世界。为了减轻这种文化或其他方面的先入为主的观念的不利影响,研究者应当逐字记录人们的话语,避免对人们的评述只记录摘要,或者以一种更加熟悉的方式来重述人们的话语。

研究者应该经常重新审视田野笔记,因为这个过程能够让田野笔记不断修正和完善。这样的审视也有助于对资料的分类和分析,这个工作最好能够与数据收集的工作同时进行。例如,田野笔记可能包含了一些分析性的备忘录【Memos】,它们能对那些更具真实和方法论意义性质的田野笔记进行补充,因而构成了初步分析的核心内容。这样的备忘录可能包括在田野调查每天工作结束后写下的总结,研究者可能在其中揭示已经显露的研究主题或发展相关概念,以及有关分析框架的初步想法。不仅如此,重新审视田野笔记还可以为需要补充收集数据的领域提供指引。如果我们没有系统地考虑田野笔记对形成分析的重要价值,而任其自发地随意积累,那么田野笔记便会逐渐失

去作用，甚至变得完全无用。

记录和组织田野笔记是一项个体活动，所以研究者需要找到一种适合自己品位和需要的方式。总体而言，田野笔记包含了以下可能会有所变化的内容：每一组笔记开头都要记录日期、时间、观察地点和笔记记录时间；在适当的地方，会用人名和地名的假名；将对于人物、时间和情境的逐字描述，与那些解释性文字区分开来。田野笔记的目标是尽可能准确、快速地记录信息，因此不必过多在意语法和拼写问题。这种文字记录，应用不同的文体来区分逐字描述所记录的文字和那些基于合理和大致的回忆所记录的文字。

对田野笔记写作的精通是通过长时间不断实践来获得的。但长时间的实践依然无法改变一个事实，那就是写作田野笔记的过程是枯燥乏味的，它需要耐心和坚持不懈。尽管如此，这项工作并非只是一项苦差事：写作田野笔记经常伴随着灵感的闪现、情绪的兴奋和对调查的社会世界的理解。

大多数观察者都依靠传统的记忆方式来记录资料，但现在研究者使用记录设备的情况也并不少见。尽管没有必须使用设备进行记录的规定，但大家已经有了基本共识，即研究者的主观性会损害收集资料的质量，应当在深思熟虑后合理地使用记录设备，当然设备的使用也必须是在研究参与者对场景及其设备较为熟悉之后。

在计算机程序出现之前，把田野笔记进行归类的任务是十分艰难的。现在针对定性资料的计算机程序可以完成很多枯燥、费时和机械化的操作处理，既包含了传统的分析程序，也允许对田野资料进行快速检索。这些计算机操作可以极大地增加效率和精确性，使研究者能够投入更多精力进入资料分析的解释阶段。但需要注意的是，如果没有研究者的概念输入，程序和技术本身对田野笔记的组织和分析所起的作用是微乎其微的。

一个好的民族志研究必须建立在田野笔记足够好的基础上。令人费解的是，业已出版发表的关于田野研究方法的著述，对如何写作和组织田野笔记这一问题着墨不多。它们的重要性被人们忽视，因为大家都普遍认为，如果田野研究者无法及时做田野笔记，他们或许也不应该在研究情境中花费过多时间。

——William Shaffir

（郑晓娟译　沈崇麟校）

## 参考文献


Berg, B. L. (2001). *Qualitative research methods for the social sciences.* Boston: Allyn and Bacon.

Lofland, J., & Lofland, L. H. (1995). *Analyzing social settings: A guide to qualitative observation and analysis.* Belmont, CA: Wadsworth.

Taylor, S. J., & Bogdan, R. (1998). *Introduction to qualitative research methods: A guide & resource.* New York: Wiley.


# 抽屉问题（File Drawer Problem）

抽屉问题代表一种对元分析【Meta-Analysis】或研究综合结果和结论的效度【Validity】的威胁。这一现象也称为发表偏倚（publication bias），有时或也称为检索偏倚（retrieval bias），它反映对元分析一般都不能找回和包括在感兴趣的专题上进行的研究的整个总体【Population】这一事实的担心。这一名词系罗伯特·罗森塔尔

(Rosenthal,1979)所创,指那些一直藏在研究者的抽屉里的研究结果可能不同于那些包括在元分析中的结果,因而可能会使元分析的结果无效。

抽屉问题是以存在着一种选择偏倚【Selection Bias】这样一种认识为基础的,这种认识认为,在进行存在选择偏倚的研究时,一个高估效应量的研究比一个低估效应量的研究更可能进入元分析。这是因为,对任何给定的样本量,那些高估治疗【Treatment】效应的研究更可能有一个显著的 $p$ 值【$p$ Value】,因而更有可能发表。因为已经发表的研究比未发表的研究更容易查找和检索,这样就可能使元分析主要以已发表的研究为依据,因而可能导致在总体上不代表研究的实体。

已经开发出来帮助我们在做元分析时处理发表偏倚的技术有三种。其中第一组技术是设计用来探测发表偏倚的。它的最好例证是采用漏斗图的图形诊断法(Light & Pillemer, 1984)。这种方法也包括马蒂亚斯·埃格和他的同事提出的明确的统计检验(Egger, Smith, Schneider & Minder, 1997)。第二组技术通常称为抽屉分析,因为它们都涉及用零效应输入的效应量【Effect Size】的估计值数(对应于留在研究者抽屉中的未发表的研究),它需要将观察的元分析结果降为零,或有时降为理论或临床上无意义的水平(Orwin, 1983)。这些做法都以罗森塔尔(Rosenthal,1979)的原建议为依据。这一建议是计算将元分析结果降为统计上不显著所必需的有着零效应的"丢失"研究数。这一数目后来被人称为容错 $N$(fail-safe $N$)。在所需"丢失"研究数难以置信的大的时候,元分析的发现可以被认为对潜在的发表偏倚效应是稳健的。基于少数研究的元分析尤其容易受到抽屉问题的侵害。在这种情形中,少数未曾检索的有着与那些已经检索的研究不同的结果的研究的检索与引进,可能会减弱,甚至推翻元分析的发现。

第三组技术被设计用来修正在某种发表选择的显式模型下,可能的发表偏倚效应的效应量估计值。这些技术中,最为著名的是休·杜瓦尔和理查德·特威蒂(Duvall & Tweedie,2000)开发的齐整和填充法(trim-and-fill method)。

可以预期,抽屉问题可能会因为更全面地将那些由研究者自己登记的和由于编辑政策的改变而将未曾发表的研究包括进来而有所改善。

——Hannah R. Rothstein

(沈崇麟译校)

## 参考文献

Duvall, S., & Tweedie, R.(2000). Trim and fill: A simple funnel plot based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, *56*, 276-284.

Egger, M., Smith, G., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal*, *315*, 629-634.

Light, R., & Pillemer, D.(1984). *Summing up: The science of reviewing research.* Cambridge, MA: Harvard University Press.

Orwin, R. (1983). A fail-safe N for effect size in meta-analysis. *Journal of Educational Statistics*, *8*, 157-159.

Rosenthal, R.(1979). The "file drawer problem" and tolerance for null results. *Psychological Bulletin*, *86*, 638-641.

## 研究中的影像(Film and Video in Research)

自20世纪初开始,电影(近年来则是视频)在整个社会科学中都被用于图像记录和表现人类的行为和文化。电影首先是在19世纪末用于定性研究【Qualitative Research】的,比如说,当时的人类学家哈登就曾经用电影拍摄了托雷斯海峡的探险。人类学家制作作为档案文献片段或文化的影像表现的电影,发展出了人类学电影这个纪录片类型。罗伯特·弗莱厄蒂(Flaherty,1921)的电影《北方的那努克》(*Nanook of the North*)是对北美爱斯基摩人的生动描写,它被称为第一部民族志电影。今天民族志电影经常采用一种观察电影的风格。比如20世纪后半叶,格拉纳达电视台制作的《正在消失的世界》(*Disappearing World*)系列电视片。近年来,数码录像技术的发展提高了影像质量、降低了编辑过程的成本,把民族志电影人吸引到录制视频而不是拍电影上。比如说,大卫·麦克杜格尔(MacDougall,2001)就提出,相比符合电视台播放要求的高成本电影来说,数码录像让民族志学者得以制作更符合学术原则的电影。麦克杜格尔讨论了他在印度的研究中是如何运用数码录像来更深入地探讨新的议题和问题,并更为紧密地与他的电影主题配合。

把电影和视频作为研究手段的做法贯穿了整个20世纪。电影和视频仍用于对焦点小组【Focus Groups】等研究活动的视频记录,还用于研究肢体动作和姿势。这包括将电影的使用延伸到参与式观察【Participant Observation】和深度访谈【In-Depth Interview】中(这两者都已是民族志电影制作的组成部分了)。例如,在对一个舞蹈队的研究中,托马斯(Thomas,1997)通过为排练录像的方式参与其中。随后,他的录像也会放给舞蹈队看。巴恩斯、泰勒-布朗和韦纳(1997)录影了感染HIV的母亲讲给子女的话。和托马斯一样,巴恩斯等人制作的视频材料不仅对其社会学分析有用处,而且对其研究对象也有用处。这个案例制作的视频将来就会放给这些妇女的孩子们看。同样,研究者也可以要求知情人以视频日记来记录其日常生活、经验、身份认同等各个方面。这些方法让研究者得以制作一个访谈或研究情景的视频和音频记录以及一个描绘研究经验的视频文本,因此,可用它来探查研究中知识生产的过程。

自1990年代晚期以来,数码录像技术的发展让影像研究方法在民族志研究中迅速普及开来。事实上,数码录像提供的可能性意味着它将会成为一个大受欢迎的研究工具。新的数码技术也能实现将视频直接从相机导入个人电脑。这对民族志电影的视频编辑、视频或其他研究材料的数码分析和保存都带来了很大的便利。因此,当代的影像研究者只需要有限的技术就可以相对独立地开展工作。

——Sarah Pink

(杨可译 沈崇麟校)

### 参考文献


Barnes, D. B., Taylor-Brown, S., & Weiner, L. (1997). "I didn't leave y'all on purpose": HIV-infected mothers' videotaped legacies for their children. *Qualitative Sociology*, *20*(1),7-32.

MacDougall, D. (2001). Renewing ethnographic film: Is digital video changing the genre? *Anthropology Today*, *17*(3),15-21.

Pink, S. (2001). *Doing visual ethnography: Images*,

*media and representation in research*. London: Sage.
Thomas, H. (1997). Dancing: Representation and difference. In J. McGuigan (Ed.), *Cultural methodologies*. London: Sage.

## 一阶(First-Order)

参见阶【Order】。

## 钓鱼式考察(Fishing Expedition)

钓鱼式考察描述的是一种研究者对即将进行的研究尚无具体目标的这样一种情况。一般来说,这个术语可能指的是(通常是负面意义的)研究者没有根据早前的研究形成假设,而只是像"审前调查"一样寻求要讨论的结果。更具体地说,它也可能用于描述在缺乏特定的指导性假设的情况下就使用统计数据,以期找到统计上显著的关系。因为如果采用 0.05 的显著度,那么就有 1/20 的可能发现统计显著的结果。

——Tim Futing Liao

(杨可译　沈崇麟校)

## 固定效应模型(Fixed Effects Model)

方差分析【Analysis of Variance, ANOVA】模型可以用来研究因变量是如何取决于一个或几个因素的。这里的一个因素定义为一个分类的自变量;换句话说,是一个定类测量层次【Level of Measuremint】的解释变量。一个固定效应的 ANOVA 模型(也称作第一类方差分析模型)有固定的(即非随机的)参数来表达每个因素的类别效应。如果统计推论的目标在发现那些完全适用于出现在手头数据中的因素的分类的结论,那么这种模型就是合适的。这和随机效应模型【Random Effects Models】不同,随机效应模型在另一词条中有介绍,如果因素类别的观察是来自一个总体的样本,那么随机效应模型就是合适的。例如,固定效应 ANOVA 模型可以用于研究雇员的满意度如何取决于性别和类别数不多的工作类型。

固定效应 ANOVA 模型是广义线性模型【General Linear Model】的特殊情况,而广义线性模型是回归【Regression】分析的基础。在 ANOVA 中一个因素各个分类的效应可以用回归分析中虚拟变量的回归系数来表示。广义线性模型可以写成

$$Y_i = b_0 + b_1 X_{1i} + \cdots + b_p X_{pi} + E_i$$

式中 $i$ 表示这个案例,$Y$ 表示因变量,$X_1$ 到 $X_p$ 表示自变量(或解释变量),$E$ 是未解释的部分,通常称为残差或误差。数量 $b_h$ 则是变量 $X_h$ 的回归系数。回归系数是固定(即非随机)的数量,也被称为固定效应。它们是总体的特征,回归模型据此而定义。在其他更复杂的模型中出现的固定效应也是类似的,例如,广义线性模型或非线性回归模型。

无论因素的各个分类是否能被作为某个总体的一个样本,我们都认为某个给定因

素的效应按固定而非随机效应建模这种做法还有另一个原因,因为这些因素或变量虽然并非研究者主要感兴趣的,但为了达到更好的模型设定却必须将它们包含进模型,这些变量被称为扰攘变量(nuisance variables)或控制变量。在许多情况下(取决于模型的正确设定),这些扰攘变量用固定效应包含进来比用随机效应包含进来能控制得更好。"控制得更好"意味着对主要感兴趣的参数的估计偏差更小,或者完全没有偏差。有一类例子是用于计量经济学的固定样本追踪数据,在这些例子中的某些研究,时间变量(常常是年代)都在起着一种扰攘变量的作用,即把它作为一种固定效应包括进来,并对它加以控制(参见 Greene,2003)。另一个例子是此前提到的根据性别和工作分类研究雇员的工作满意度,现在少量或大量不同的公司里进行研究。公司如果被当作扰攘变量,则可以当作固定效应来处理。另一方面,如果对公司间的变异性的量和类别感兴趣,公司也可以作为随机效应纳入模型,这样就形成了混合效应模型【mixed-effects model】。

——Tom A. B. Snijders

(杨可译　沈崇麟校)

## 参考文献


Cobb, G. W. (1998). *Introduction to design and analysis of experiments.* New York: Springer-Verlag.

Greene, W. (2003). *Econometric analysis* (5th ed.). Upper Saddle River, NJ: Prentice Hall.

Iversen, G. R., & Norpoth, H. (1987). *Analysis of variance* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-001, 2nd ed.). Newbury Park, CA: Sage.

Neter, J., Kutner, M., Nachtsheim, C., & Wasserman, W. W. (1996). *Applied linear regression models.* New York: Irwin.


# 地板效应(Floor Effect)

如果某个测量对可能的回答有一个明显的下限,而大量参与者的分数都集中在或接近这一下限时,就出现了地板效应(与天花板效应【Ceiling Effect】相对)。每当方差受制于这种方式时,就会出现一种称作尺度衰减的方法论问题。这个问题通常在实验研究环境中来讨论,但它可能会威胁到任何类型的研究的效度,让其测量指标在可能的取值范围的下限几乎体现不出任何变化。

许多因素都可能会引发地板效应。如果结果量度涉及一个有下限的绩效任务(例如正确回答的数目),参与者可能会觉得认为太难了,从而无法给出正确回答。评分量表也有下限,而参与者的回答可能会落在反应量表的取值下限之外。

因为缺少方差,地板效应对实验研究、非实验研究的效度都是一个严重的威胁,任何有地板效应的研究都应谨慎地加以解释。如果因变量遭到地板效应影响,实验或半实验操作就会表现得没有效果。例如,评定记忆的一个普通方法就是能回忆起来的条目占多少百分比。如果记忆任务太难,那么所有的参与者可能都无法回忆起任何条目。在有关回忆的感觉形态效应的研究中,如果地板效应实际上无法使我们发现差异,那么研究就可能会无功而返。

地板效应在解释交互效应【Interaction Effects】时尤其麻烦。许多交互效应都是无法在某个自变量的一个水平上发现明显差别,而在另一个水平上却有明显差别。无法发现明显差别可能是由地板效应导致的,对这样一种交互效应的解释应当十分谨慎。

地板效应对非实验或相关研究也是一种威胁。许多推论统计都以相关变量正态

分布的假定为依据。它可能使我们无法计算显著性检验,或者受限制的方差范围可能会在虚无假设实际上是真的时候,增加拒绝它的可能性(第二类错误【Type Ⅱ Error】)。

解决地板效应问题的最佳方案是预检验,它使我们可以较早地认定问题。如果发现了地板效应,可以用几种方法来处理。如果结果量度是任务完成,那么我们任务可以弄得简单一点,以扩大潜在回答的取值范围。对于评定量表,潜在回答的数量可以扩大些。其他的处理方案包括在评定量表中使用极端参照刺激等。在评定感兴趣的变量之前,可以要求参与者评价一个刺激物,这样他们就会用到分布中的极端值。这样会鼓励参与者使用评定量表中较为中间范围的取值,减少地板效应的可能性。

——Robert M. Hessling
Tara J. Schmidt
Nicole M. Traxel
(杨可译 沈崇麟校)

## 焦点小组(Focus Group)

焦点小组是一种研究访谈的过程,专门设计出来用于在一小群被调查者中发现深刻的见解。小组访谈的独特之处在于,它采用的一系列问题的顺序都是精心编排过的,或者专注于将讨论导向研究者关注的概念上。组成焦点小组的参与者人数有限且是同质的,在一个宽松的、没什么威胁的环境里讨论事先拟订的题目。由一位有经验的主持人员或访问员来指导讨论,对讨论进行有限的控制,并令它从一个问题转向另一个问题。

尽管这被称为焦点小组访谈,但其实更像聚焦讨论。主持人员要使参与者参与交谈,并鼓励他们对组内其他人的谈话进行评论。主持人员的主要责任就在于保证讨论的切题并控制时间,并且让参与者积极参与交谈。

焦点小组访谈在许多方面都有应用受人欢迎,因为这个程序有助于研究者理解其目标受众的行为、习惯和深刻思想。这个信息被用于理解消费者对新的节目和产品会有什么反应,以及哪些行为会影响消费者行为。近年来,焦点小组从消费者研究转移到了学术研究、项目评估、社会化营销、需求评估以及有关的研究上。焦点小组作为对其他方法的补充和改进,在研究中用得越来越多,比如说,对抽样调查工具的发展或在寻找替代性解释方面都发挥着作用。

### 焦点小组访谈的发展

"焦点小组访谈"这个术语源于第二次世界大战前罗伯特·默顿的研究。哥伦比亚大学的社会学家罗伯特·默顿和保罗·拉扎斯菲尔德引入了小组访谈,这种方法有一个焦点明确的提问顺序,请人对行为、态度和观点发表评论。罗伯特·默顿与肯德尔和玛乔丽·菲斯克(Merton, Fiske, & Kendall, 1956/1990)一起合著了《焦点访谈》(*The Focused Interview*),这是一本有关宣传材料和军队士气的书。

焦点小组在 1950 年代晚期受到了青睐,当时市场研究者正在寻求研究消费者购买行为的办法。从 1950—1980 年代,焦点小组大多数时候被用于寻找消费者的心理动机。许多著名的焦点小组实施者接受的都是心理学训练,其目的是发现对消费者行为的无意识联系。那时,焦点小组由 10~12 名参与者组成,通常都彼此陌生,在单向镜

后进行，使讨论发起人有机会进行观察。

自从 1980 年代以后，学者也开始把焦点小组当作一类定性研究。学者们以罗伯特·默顿和同事的工作为基础，修改了市场研究者所采用的程序。这些较晚时期的小组更为强调在自然的环境中开展研究（如家里、公共的聚会空间），参与者更少（通常是 6~8 人），在分析中也更为严格（系统的、可检验的程序）。

焦点小组研究是一个不断演变的研究领域。多年来，研究者们对这些技术进行了调整和改变，以适应具体的环境。电话或网络焦点小组作为对受距离阻隔的人们展开调查的一种办法，日益受到欢迎。在国际化的环境中，读写能力可能会妨碍其他形式的调查，但焦点小组却表现出相当好的通用性。

## 焦点小组研究的质量

随着焦点小组日益深受欢迎，其质量也开始受影响。对一个随意的观察者来说，焦点小组可能会很有迷惑性，看起来很简单，但实际上它在策划、开展和分析上都很复杂。当研究者没有足够的准备、计划和技巧就进行焦点小组研究时，其结果就不太可靠。焦点小组研究的质量受四个关键因素的影响：

• 策划，包括确定合适的受众，让这些人聚到一起来。

• 询问正确的问题。

• 在避免把谈话引入预定答案的前提下，主持讨论。

• 在准备报告时采用系统的、可检验的分析策略。

策划焦点小组研究包括决定小组的人数、找到招募小组成员的办法、制造刺激、预先考虑其他的后勤问题。基本的研究策略就是要为研究者感兴趣的每种受众类型进行 3~4 个焦点小组研究。在决定样本规模的问题上，统计公式并不合适。相反，研究者采用一种称为冗余或理论饱和【Theoretical Saturation】的概念。研究者用冗余或理论饱和的概念一直继续访谈，直到不再出现新的洞见为止。实际上，研究者已经对这个主题的各种观点的范围大致有了解。继续访谈只会产生更多研究者已经知道的内容。在焦点小组研究中，这大概会在一种受众之中进行 3~4 组焦点小组访谈之后出现。

招募从尽可能准确地确定目标受众的特征开始。参与者被选中、得到邀请，是因为他们具有共同的某种经验或品质。邀请的是满足这些“筛子”或研究资格的人。

成功的招募有三个明显的特征：首先，这个过程应该是个人化的。这意味着每个受邀的人都感觉是他或她本人被邀请参加和分享观点。其次，邀请过程是反复的。邀请不止是一次，而应该是两次、三次。最后，要有鼓励参与者的奖励措施。奖励可以是钱、食物、礼物或无形资产，例如，为一项有价值的研究提供帮助、帮助某个邻里组织，或者为这位参与者投身其中的事业提供支持。各个研究的奖励措施可以不同，因为参与者的兴趣各有不同。

必须为焦点小组准备好问题。这些问题应该设计成聊天式的，便于参与者理解。在一个两小时的焦点小组访谈中，可能会有十几个问题。在问题措辞和排序上要小心。焦点小组的问题与众不同之处就在于其焦点，对于具体经验的强调，以及引发谈话的倾向。

这是在设计问题时经常考虑的一些因素：

• 采用开放式问题【Open-Ended Question】，让参与者在回答问题时的自主度越大越好。

• 避免采用能以“是”或“否”来回答的

问题,因为这些回答无法提供研究通常需要的细节解释。

• 不要去问“为什么”,因为“为什么”听起来像是质问,让人们变得有戒心。相反,要去问特点和/或影响。

• 用“回想式”的问题来让人们回到某个具体时点,以获取基于真实经验的信息。

• 用能让参与者投入的问题来提问,例如比较、评价、分类、描绘等。

• 聚焦在问题上,按从一般到特殊的顺序来排序,这样问题就会自然而然地过渡到最重要的主题上来了。

• 用封闭式问题来结束小组访谈并从参与者那里获得概括性评价。

主持是指导讨论的过程。一开始,主持就是要帮助人们感觉到十分舒服地在小组中和他人分享自己的所思和所感。主持人员要建立一种相互信任的、宽松和舒服的氛围,扫除交流的障碍。主持人员要介绍这项研究,帮助参与者互相熟悉起来。在焦点小组访谈一开始,主持人员要描述讨论的协议【Protocol】,说明谈话会被录音。要给予参与者保密的承诺,告诉他们尽管可能会引用他们的话,但在后来的报告中不会出现他们的名字。给讨论建立好几条基本规则之后,主持人员就开始问问题。主持人员在谈话进行时要有几个关注点。主持人员必须知道什么时候要等待更多的信息,什么时候要转向下一个问题。主持人员必须能够控制强势的讲话者,鼓励犹豫的参与者。主持人员还必须尊重参与者,倾听他们必须要说的话,感谢他们的观点,即便他或她本人可能并不同意。主持人员引导参与者完成所有问题,小心控制时间,当得到足够的回答之后就从一个问题转向另一个问题。

分析就是让一连串会话变得有意义的过程。不是所有的研究都要求进行同样水平的分析。有时候,分析可以做得很快,因为数据很清楚,模式很明显。但有时候分析过程可能会特别耗时。分析过程包括数据评析和寻找各小组共同的主线主题。因为数据是由多种方法获取的,包括记忆、田野笔记、录音和/或视频录像等,所以焦点小组访谈的通常都需要进行转录,而这些转录的文字则会用于后来的分析。

焦点小组分析应该是系统的、可检验的。系统的分析意味着分析者所起草的方案必须有一个预先确定的捕捉和处理数据的顺序。而这个过程是可检验的则是指它是有据可查和可以重复的。

## 什么时候使用焦点小组

同所有的研究方法一样,有些情况下焦点小组会特别有效,有些情况下应该用别的方法。焦点小组在以下情况中特别奏效:

• 寻找是什么激起了某种行为,例如购买、使用或采取行动。

• 寻找阻碍某种行为的障碍是什么。

• 预检验理念、竞选、抽样调查、产品等。在计划开始实施、大量的钱被花出去之前,焦点小组可以用于获取对计划的反馈。

• 理解人们是如何考虑或感受某个事物的,例如某种观念、行为、产品或服务。

• 开发其他的研究工具,例如抽样调查或个案研究【Case Study】。

• 评估项目运行得怎么样,产品工作情况如何,以及如何改进它们。

• 理解人们如何看待他们生活和社区中的需要和财产。

相反,比较起来,焦点小组访谈并不适用于以下几种情况:

• 一个让人们达到共识的过程。

• 一种教授知识或技能的方法。

• 一种知识或技能的测试。

在以下场合也不宜使用焦点小组：

• 在对某个总体做统计推论时。焦点小组研究涉及的数量太小，不足以做统计推算。

• 主题太有争议性，参与者想要争论而不是讨论。

• 没有培育信任和尊重的环境。如果赞助方或机构被认为是不值得信赖的，或者如果社区发生的事情让人们犹豫不决或愤世嫉俗，那么这个研究结果就会受损。

• 话题不适合小组讨论，适合个别讨论。

请注意：尽管研究者承诺了保密性，实际上，他或她无法保证参与者会一直保密。

## 焦点小组访谈的方法

焦点小组访谈是一个正在不断改进的方法论，以适应不同的情况和环境。下面要讨论的就是一些最常见的形式。

### 市场研究焦点小组法

在这些小组中，专业的主持人员带领着10~12名同质的领取报酬的参与者，就有关某个消费品、服务或组织认同的主题开展讨论。这些小组讨论通常都在装有单向镜的特殊的焦点小组室进行，围绕着一张圆桌，大约2小时。结果和报告很少跟参与者分享。参与者一般不了解赞助组织的身份。通常在数周之内就需要得到结果。

### 学术性的焦点小组法

在这些小组中，主持人员是一位学术性的研究者，他或她带领着由6~8位同质的参与者组成的小组，就某个具有学术旨趣的主题进行讨论，这个主题可能会为某个理论概念或参与者对现实的理解提供洞见。这些小组讨论通常都是在方便、舒适的社区环境中而非装有单向镜的特殊房间里进行的。参与者可能会得到经济上的奖励。结果后来会发表在学术性杂志或论文中。对于仔细的数据收集和系统的分析过程也非常看重。这些研究通常都要进行6~12个月。

### 公共/非营利焦点小组法

在这些小组中，主持人员可能是一个内部员工、一个志愿者或是一个局外的顾问。这些小组比市场研究法的小组更小（通常是6~8位参与者）。这些小组访谈在社区环境中进行，通常会避免使用单向镜，以节省租金费用，避免被观察者产生敏感。研究设计的主题是为了改进服务或产品、评估项目或政策、理解消费者满意度，或者对公共的、宗教的、教育的、服务组织的其他需求进行评估。研究结果几乎总是要和社区成员进行分享。这些研究通常在几个月的时间内完成。

### 参与式焦点小组法

这些焦点小组由志愿者、社区成员或那些愿意作为一个研究团队在一起工作的可能的工作人员领导。通常有一个专业的研究者作为导师和教练，帮助这个团队进行准备、设计问题、承担复杂的分析任务。志愿研究者进入他们的社区、招募参与者、用预先设定的问题来进行焦点小组访谈。这些研究在获取有关预防工作、健康与安全问题，以及社区需求评估方面的洞见上很有效。这些小组主要的优势就是志愿者在社区里关系很融洽，可以很容易在谈话中建立信任、吐露真话。这些结果都要反馈给社区成员共享，通常需要几个月的时间来收集。

### 电话焦点小组法

电话焦点小组与电话会议差不多。它由一个主持人员主导，包括同时电话连线的4~6名参与者。主要的优势就是便捷。研究者可以用很少的费用，对散布在全国的一

组人进行研究。通常会在小组访谈之前把较少的几个问题(5～8个)发给参与者。这些小组访谈通常会持续1小时。

儿童焦点小组

这些小组进行了改造,以适应儿童的兴趣和能力。考虑到发育的差异,参与者年龄都很相近(通常年龄差不超过两岁)。参与者年龄可以低至10岁,但他们必须能够倾听小组的其他人讲话进行交谈。小组的规模较小(5～7人),小组访谈的时间可以短一点。儿童小组的问题通常比成年人小组包括更多的活动和运动。

互联网焦点小组和聊天室

这些小组在不断演变,并且已经有了大量的变化形式。有些小组有点像聊天热线,少数参与者登录同一个聊天室中心。主持人员在线提出问题,参与者以书面形式提供答案。还有一种变化形式,就是参与者登录一个特别的网站,在一连数天内每天都会得到不同的问题。问题通常是互倚的,参与者也可以看到其他参与者的评论。随着技术的进步,这些电子焦点小组一定会有更多的选择。

——Richard A. Krueger

(杨可译　沈崇麟校)

## 参考文献

Krueger, R. A., & Casey, M. A. (2000). *Focus groups: A practical guide for applied research* (3rd ed.). Thousand Oaks, CA: Sage. [This one-volume (215 page) overview of focus group interviewing provides a description of what focus groups are, how they are used, and strategies that make them successful.]

Merton, R. K., Fiske, M., & Kendall, P. L. (1990). *The focused interview* (2nd ed.). New York: Free Press. (Originally published in 1956, Merton and his colleagues describe the development of the focused interview.)

Morgan, D. L., & Krueger, R. A. (Eds.). (1998). *Focus group kit*. Thousand Oaks, CA: Sage. (A collection of six books, each approximately 100-140 pages, on focus group interviewing. Specific volumes include: *The focus group guidebook*, *Planning focus groups*, *Developing questions for focus groups*, *Moderating focus groups*, *Involving community members in focus groups*, and *Analyzing and reporting focus group results*.)

# 聚焦比较(Focused Comparisons)

参见结构的聚焦比较【Structured, Focused Comparison】。

# 聚焦访谈(Focused Interviews)

这种访谈形式最初设计出来是用于个人和小组的(Merton, 1987; Merton & Kendall, 1946),但被大多数人都认为是焦点小组【Focus Group】法。聚焦访谈有一个明确的目标,即在被调查者亲历的某个特殊的具体情境下(例如,一个共同经历的事件或非常特别的经历)来测试、评价或生成假设。访谈的焦点由相关理论和证据圈定(在一对一的讨论或小组讨论中),还需要熟练运用访谈提纲【Interview Guide】来把过程加以简化,也要给挖掘未曾预料到的观点留出余地。合理利用这一方法关系如何把握一

个悖论,即在以自由流畅的讨论探求真正的主观信息和方法论的严格需求之间保持平衡。

聚焦访谈的目的是要超越访谈对象对其经验的概括性判断(如不愉快的、刺激的),来"准确地发现是什么深层的情感在发挥作用"(Merton & Kendall, 1946, p. 541)。因此,这个技术的关键就是熟练地运用语言刺激来引发对问题的具体回答,而不是模糊概述,同时还要具有不断检测和调整访谈过程的能力,以便让访谈对象尽可能地自我表露。在用于小组时,这个过程给我们带来了一定的便利性,也给我们发出了一些挑战——我们应该怎么样来掌控和协调多个可能的被调查人,以保证导出有足够广度和深度的证据(Millward, 2000)。在此情况下,重要的设计要素包括小组规模(例如,既要大到能抽出足够广泛的观点,又不会有什么禁忌)、小组构成(如教育的同质性)以及空间安排(如坐成一圈)(Merton,1987)。

聚焦面对面访谈现在是方法论知识库中的重要部分,尤其是在临床或健康研究中。但直到1980年代末,聚焦小组访谈的研究优势才在其市场领域的应用之外正式得到社会科学家承认(Morgan,1988)。两种类型的聚焦访谈都可以作为定性数据收集的主要手段(如为了构建假设),或作为其他方法的附属(如为阐释研究发现提供帮助)。理论也可以当作聚焦工具来提供构念,围绕着构念可以展开一对一或者小组对话(例如保护动机理论),尤其是在涉及复杂或敏感议题时(如对青少年风险性行为的检查)。

如果被访谈人同意,访问的回答可以被录音或录像,并转录成文字,为某种形式的定性数据分析做准备。在最基本的层次上,内容分析【Content Analysis】就涉及内容的类别。其他用于聚焦访谈的更具体的分析形式包括阐释性现象学分析和话语分析【Discourse Analysis】等。话语分析既考察讨论的过程,也要考察其内容。最终分析策略的选择还是要取决于认识论和理论因素的结合,包括确切说明提出的问题究竟是什么(Millward, 2000)。

——Lynne Millward

(杨可译　沈崇麟校)

## 参考文献

Greenbaum, T. L. (1998). *The handbook for focus group research*. London: Sage.

Merton, R. K. (1987). The focused interview and focus groups: Continuities and discontinuities. *Public Opinion Quarterly*, *51*, 550-566.

Merton, R. K., Fiske, M., & Kendall, P. L. (1990). *The focused interview: A manual of problems and procedures* (2nd ed.). New York: Free Press.

Merton, R. K., & Kendall, P. L. (1946). The focused interview. *American Journal of Sociology*, *51*(6), 514-557.

Millward, L. J. (2000). Focus groups. In G. M. Breakwell, C. Fife-Schaw, & S. Hammond (Eds.), *Research methods in psychology* (2nd ed., pp. 303-324). London: Sage.

Morgan, D. L. (1988). *Focus groups as qualitative research*. Newbury Park, CA: Sage.

Morgan, D. L. (1993). *Successful focus groups: Advancing the state of the art*. London: Sage.

Stewart, D. W., & Shamdasani, P. N. (1990). *Focus groups: Theory and practice* (Applied Social Research Methods Series, Vol. 20). Newbury Park, CA: Sage.

## 强迫选择测验(Forced-Choice Testing)

在强迫选择测验中,被测验者必须从备择的答案中进行选择,它与被测验者自行提供他们自己构建的答案的自由答案测验有所区别。在社会科学中,如何在这两种形式的测验中进行选择,一般取决于测验的目的,强迫选择测验适用于对一个内容广泛的领域迅速地作出评定,使得数据很容易在数量上作出解释。自由答案测验则针对答案进行定性的分析,因而一般需要花费更多的时间对答案进行解释。

强迫选择测验有各种各样的形式。多种选择测验给被测验者以两个或更多个可供选择的答案。双备择答案的强迫选择测验包括"真-假""是-否"或"正确-不正确"这样的答案对。多种选择的能力测验,包括一个正确的答案,而其余答案则都是备择的不正确的答案。态度量度包括各种形式的李克特【Likert】类量表。在这种量度中,回答人从中选择一个与陈述一致的量表。这种量表一般是一个五级的,"从完全同意"到"完全不同意",在中心点的选项一般都是"中性的"或"悬而未定的"。Q 排序技术对态度和自我概念量度是很有用的。将个人态度陈述或人格特征按他们自身的特征从最少的特征到最多的特征排列分组。一般来说,组的分布都是强迫的,以防止答案的偏斜。

人格测验已经采用对回答的一致性(相同的措辞的题项是否以相同的方式作答),随机回答的存在("内容未回答")和精神病理学的低报或高报进行评估的策略(Greene, 2000)。在某种考察设置中(如人员选择和儿童监护评价),被测验者可能有意低报精神病理学问题。在审判前的刑事考察、人身伤害诉讼和残疾判定中,被测验者可能会有意高报精神病理学问题。将简单的计数与常模性数据进行比较,使我们很快就能对一组答案作出测定。

在人格测验中,随机回答问题可以用答案一致性量表来评估。在能力测验中,完全随机回答问题不是一个问题,除非被测验者事先觉察到被视为残疾有一定的好处。强迫选择测验使我们能进行贡献分析,猜测正确的答案的总数。已知题项真实数目的一个好的估计值可以这样来计算[(正确数)-(不正确数)]/($n$-1),$n$ 是在两个或两个以上的备择答案中选择的备择答案数(Cronbach, 1990)。

强迫测验使我们能对认知和知觉测验中的捏造问题进行系统的分析。潘克拉茨(Pankratz, 1979)普及了将正确答案数与随机回答的期望数进行比较的技术,给"低于可能性"的打分指出装病。弗雷德里克森和克罗斯比(Frederick & Crosby, 2000)指出,Cronbach 法能求得明显装病的真认知能力的估计值。不仅如此,弗雷德里克森和克罗斯比还证明,比照双备择答案的平均回答困难,绘制平均回答精度图,使用困难的层次结构的强迫测验,使我们能对回答,按顺从、不一致、随机或"抑制"(装病)进行分类。

——Richard I. Frederick

(沈崇麟译校)

### 参考文献

Cronbach, L. J. (1990). *Essentials of psychological testing* (5th ed.). New York: HarperCollins.

Frederick, R. I., & Crosby, R. D. (2000). Development and validation of the Validity Indicator Profile. *Law and Human Behavior*, *24*, 59-82.

Greene, R. L. (2000). *The MMPI-2: An interpretive*

*manual*. Boston: Allyn & Bacon.

Pankratz. L. (1979). Symptom validity testing and symptom retraining: Procedures for the assessment and treatment of functional sensory deficits. *Journal of Consulting and Clinical Psychology*, *47*, 409-410.

## 预测(Forecasting)

“即使是没有确定性的预见,也远比完全没有预见要好。”

——亨利·庞加莱(Poincaré,1913,p.129)《科学的基础》(*The Foundations of Science*)

预测是一种有关未知和不确定的事件的陈述——在大多数情况(并非在所有情况)中都是有关未来事件的陈述。如果事涉时间,那么预测(或预报)是一种基于在过去的事件中观察到的规律性对于未来的结果的断言。我们假定,在单个时间系列上的观察也适合连续的时间周期。观察可能是月度的销售数据,季度的GNP(国民生产总值)、年度杀人犯数或每天的股票收盘价。过去的观察 $y_1, y_2, \cdots, y_n$ 被用来计算未来的观察 $y_{n+r}$ 的预测值 $\hat{y}_n(r)$。在这里,$n$ 是预测的原点,$r$ 是预测的时间周期,$\hat{y}_n(r)$ 是从原点 $n$ 开始,$r$ 个周期的向前预测。这里的问题是在计算预测值时,究竟需要用多少个有效的观察。在这里,我们假定不存在其他的可帮助我们进行未来的各个 $y$ 的预测系列。

指数平滑【Smoothing】和用自回归求和移动平均模型(ARIMA)预测的时间系列模型代表了两种成功的一元时间系列预测方法。我们将在这里对这两种方法进行详细介绍,并进一步指明它们的有用的扩展。

### 指数平滑法

我们可以区分三种基本的方法:无趋势的系列指数平滑法、数据展示趋势的 Holt 指数平滑法和包括趋势和季节分量的 Winter 指数平滑法。

#### 简单指数平滑法

一种预报无趋势系列的直观方法,它用历史平均数预报所有未来周期,其公式为:

$$\hat{y}_n(r) = \bar{y} = \frac{1}{n}[y_n + y_{n-1} + \cdots + y_1]$$

不考虑年龄,对观察作相等加权要求系列在时间上稳定。因为大多数数据系列都缺乏这种稳定,因此,最好在一个较短的、最近的时间窗口计算观察的平均数或考虑用一种加权的平均,根据他们的年龄来折算观察。简单指数平滑使用一种指数(几何)加权平均:

$$\hat{y}_n(r) = L_n = \alpha[y_n + (1-\alpha)y_{n-1} + (1-\alpha)^2 y_{n-2} + (1-\alpha)^3 y_{n-3} + \cdots]$$

$L_n$ 是在时间 $n$ 时的水平估计值。它变成了所有未来的观察的预测值。平滑常数 $\alpha(0 \leqslant \alpha \leqslant 1)$ 给最近的观察较大的权,给过去的观察较小的权。当 $\alpha = 1$ 时,预测值 $\hat{y}_n(r) = L_n = y_n$,使得我们只使用最后的观察。这叫作“naïve”或随机游走【Random Walk】预测。当 $\alpha$ 接近 0 时,求得的预测值 $\hat{y}_n(r) = L_n = (y_n + y_{n-1} + \cdots + y_1)/n$,在所有的观察上都放上了相等的权。

水平则根据 $L_n = \alpha y_n + (1-\alpha)L_{n-1}$ 做了更新;只有最后一个观察和前一个水平需要保存。如果有许多系列必须预报,那么这一点便是很重要的了。在系列的开端开始进行递归,用某一个起始值作为时间 1 的水平,这样的做法是很普遍的。一般来讲,我们会使用第一个观察($L_1 = y_1$)或最初几个观察的平均数。从这一起始值起,我们计算 $L_2 =$

$\alpha y_2+(1-\alpha)L_1, L_3=\alpha y_3+(1-\alpha)L_2, \cdots$，直至到达 $L_n=\alpha y_n+(1-\alpha)L_{n-1}$。最后一个水平估计值变成了所有未来的观察的预测值：$\hat{y}_n(r)=L_n$。

计算需要平滑常数的值，它是通过将单时间周期的向前预测值的平方最小化来估计的：

$$\begin{aligned}\mathrm{SSE} &= \sum_{t=2}^{n}[y_t-\hat{y}_{t-1}(1)]^2\\ &= \sum_{t=2}^{n}(y_t-L_{t-1})^2\end{aligned}$$

我们用图 1 中的艾奥瓦季度非农产业增长率来阐述简单指数平滑。数据（$n=127$；从 1948 年 2 月—1979 年 4 月）先在亚伯拉罕和莱道尔特（Abraham & Ledolter, 1983）的著作中被分析。时间系列图显示不存在明显的趋势，但表明系列的水平是随时间改变的。表 1 列出了用平滑常数 $\alpha=0.1$ 计算得到的值，迭代从 $L_1=y_1=0.50$ 开始。反复使用更新方程给出的 $L_2=(0.1)\times(2.65)+(0.9)\times(0.5)=0.715, L_3=(0.1)\times(0.97)+(0.9)\times(0.715)=0.741$，直至 $L_{127}=(0.1)\times(2.35)+(0.9)\times(2.681)=2.648$。所有未来的观察的预测值由 $\hat{y}_{127}(r)=2.648$ 给出。单时间区间的向前预测误差在第 15 列中列出：$y_2-\hat{y}_1(1)=2.65-0.50=2.15, y_3-\hat{y}_2(1)=0.97-0.715=0.255, \cdots, y_{127}-\hat{y}_{126}(1)=2.35-2.681=-0.331$；它们的平方和是 SSE = 122.61。平方和取决于平滑常数，因为不同的 $\alpha$ 意味着不同的预测值。在这一个例子中，任何不同于 0.1 的平滑常数都将会增加单时间区间的向前预测误差的和。

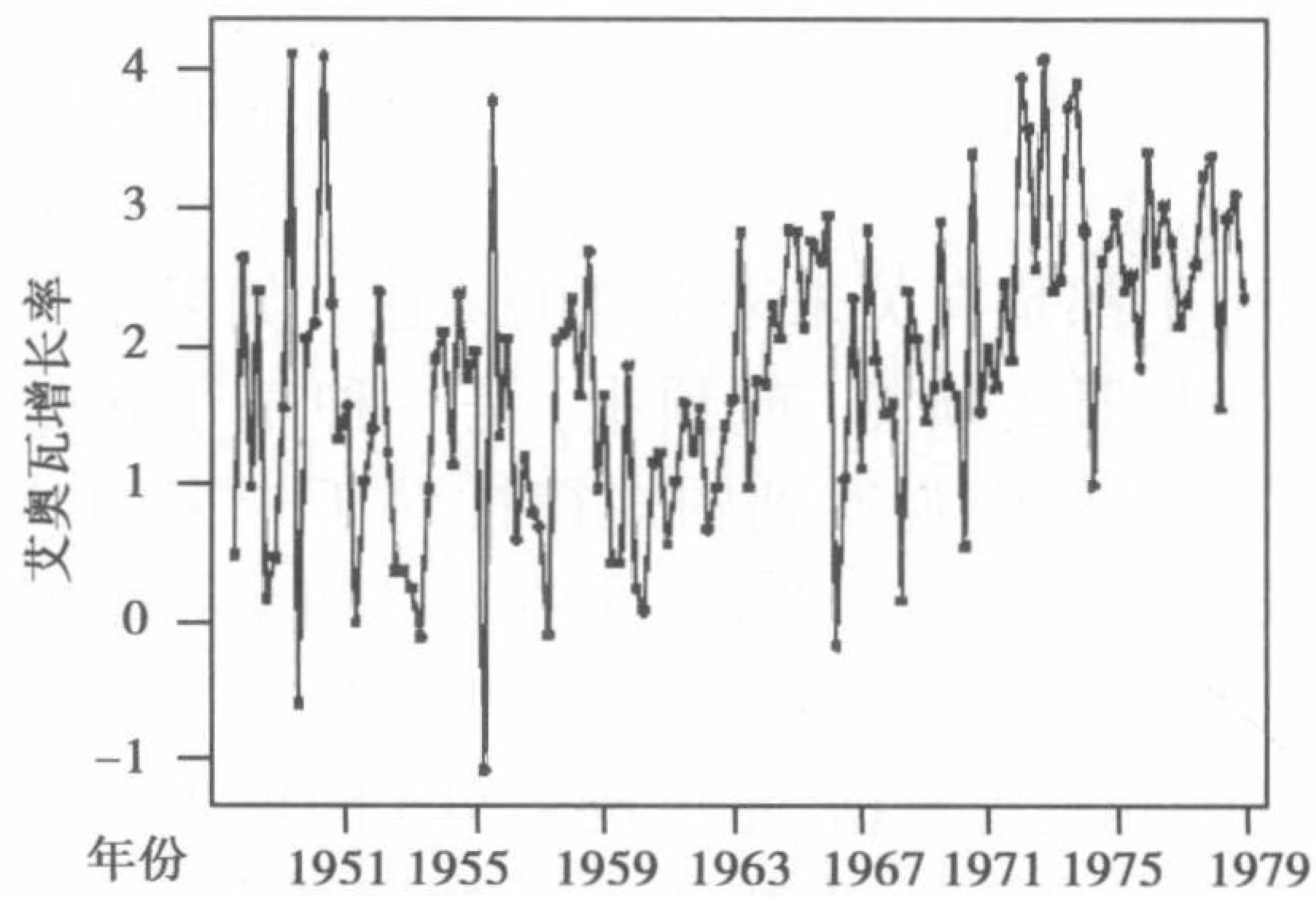



图 1　艾奥瓦季度增长率（1948—1979）

表 1　艾奥瓦增长率的简单指数平滑

| 时间 | 增长率 | 平滑值 | 向前单时间区间预测 | 单时间区间的向前预测误差 |
|---|---|---|---|---|
| 1 | $y_1=0.50$ | $L_1=0.50$ | | |
| 2 | $y_2=2.65$ | $L_2=0.715$ | $\hat{y}_1(1)=0.50$ | $y_2-\hat{y}_1(1)=2.15$ |
| 3 | $y_3=0.97$ | $L_3=0.741$ | $\hat{y}_2(1)=0.715$ | $y_3-\hat{y}_2(1)=0.255$ |
| 4 | $y_4=2.40$ | $L_4=0.906$ | $\hat{y}_3(1)=0.741$ | $y_4-\hat{y}_3(1)=1.659$ |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| 124 | $y_{124}=1.55$ | $L_{124}=2.602$ | $\hat{y}_{123}(1)=2.719$ | $y_{124}-\hat{y}_{123}(1)=-1.169$ |
| 125 | $y_{125}=2.93$ | $L_{125}=2.635$ | $\hat{y}_{124}(1)=2.602$ | $y_{125}-\hat{y}_{124}(1)=0.328$ |
| 126 | $y_{126}=3.10$ | $L_{126}=2.681$ | $\hat{y}_{125}(1)=2.635$ | $y_{126}-\hat{y}_{125}(1)=0.465$ |
| 127 | $y_{127}=2.35$ | $L_{127}=2.648$ | $\hat{y}_{126}(1)=2.681$ | $y_{127}-\hat{y}_{126}(1)=-0.331$ |

## Holt 指数平滑法

霍尔特将指数平滑法推广到了系列跟随着一种线性趋势的情形。$y_{n+r}$ 的 $r$ 个时间区间的向前预测值用下式计算：

$$\hat{y}_n(r)=L_n+rT_n$$

式中的 $L_n$ 和 $T_n$ 反映了在时间 $n$ 时的水平和趋势。我们需要这些数量的估计值，并设定更新的递归式。霍尔特认为递归式是

$$L_n = \alpha_1 y_n + (1 - \alpha_1)(L_{n-1} + T_{n-1})$$
$$T_n = \alpha_2(L_n - L_{n-1}) + (1 - \alpha_2)T_{n-1}$$

两个平滑常数 $\alpha_1$ 和 $\alpha_2$ 反映了给当前（“新的”）和以前（“旧的”）的可得到的信息的权数。在时间 $n$ 水平上的最新的信息由 $y_n$ 给出。在这个观察之前，我们有可利用的 $L_{n-1}$ 和 $T_{n-1}$，而下一个水平的最佳猜测是 $L_{n-1}+T_{n-1}$。水平的更新方程设定了这两个分量的加权平均。第二个趋势方程设定了最新趋势估计值的加权平均 $L_n-L_{n-1}$ 和前一个可利用的估计值 $T_{n-1}$。

大的平滑常数将更重的权放在了最近的观察上。平滑常数 $\alpha_1=\alpha_2=1$ 意味着 $L_n=y_n$，$T_n=y_n-y_{n-1}$ 和 $\hat{y}_n(r)=y_n+r(y_n-y_{n-1})$。所有的预测值都位于一条通过最后两个观察值的直线上。对小的平滑常数而言，观察加的权相等，与观察对时间的回归相似。

更新的计算始于合理的初始值，如 $L_2=y_2$ 和 $T_2=y_2-y_1$。递归式用于计算 $L_3$ 和 $T_3$，$L_4$ 和 $T_4$ 等，直至到达了 $L_n$ 和 $T_n$。预测值由 $\hat{y}_n(r)=L_n+rT_n$ 产生。平滑常数经由将单时间区间的向前预测误差平方和最小化求得。

在亚伯拉罕和莱道尔特（Abraham & Ledolter，1983）的著作中，将 Holt 法应用于恒温计销售的数据分析。图 2 显示 $n=52$ 的观察接下来的 8 个预测值和近似于 95% 的预测区间。Minitab 软件被用来估计平滑常数和计算预测值。

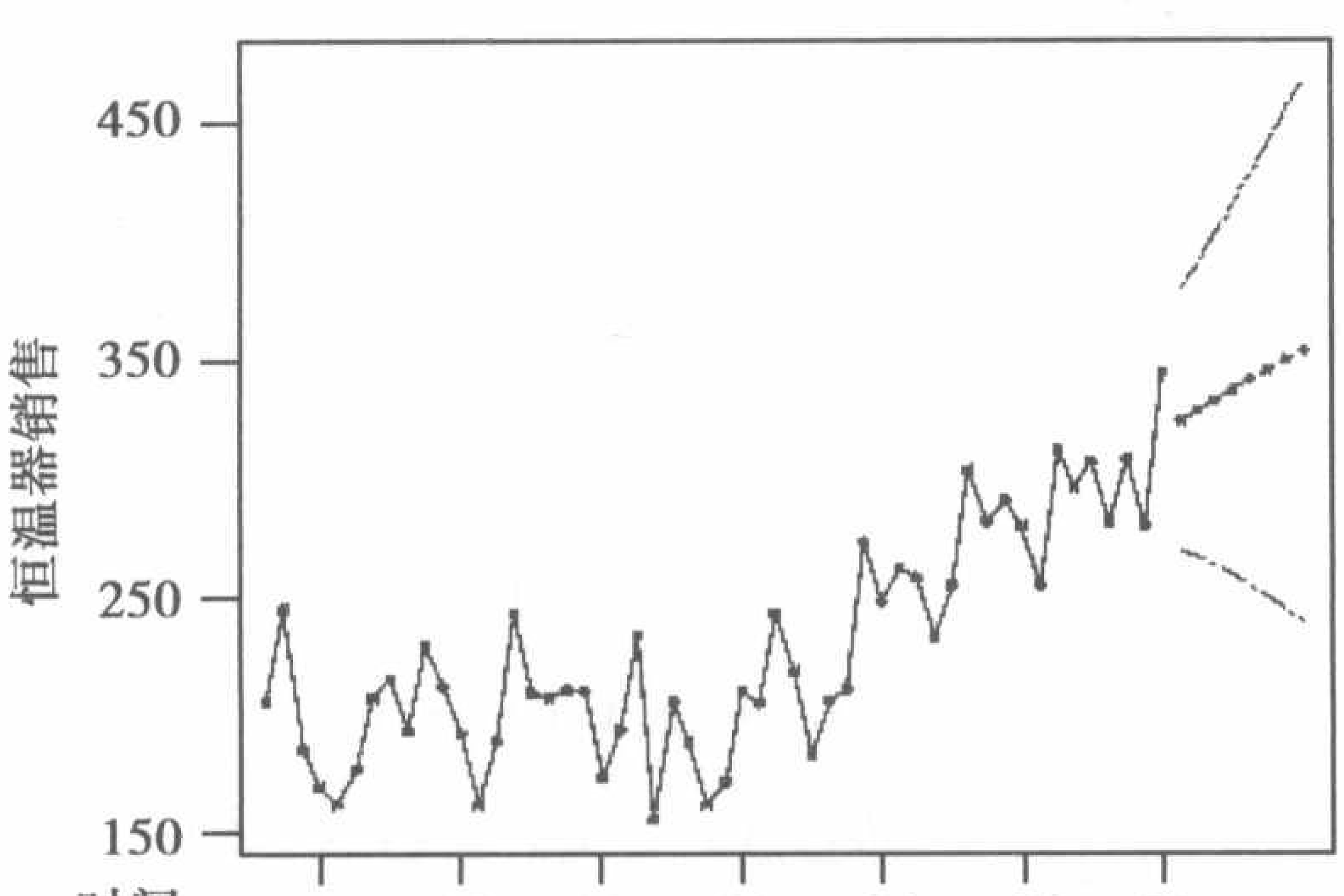



图 2　观察值、预测值和 95% 的预测区间

## Winters 季节性指数平滑法

许多系列都包括了季节性分量。销售便是典型的季节性的，在圣诞节前后销售量就很大。对季节性的数据而言，我们将 Holt 法扩展，使之带有一个额外的方程和一个第三个平滑常数。假定月度数据有一个年度的季节性分量，预测值由下式给出：

$$\begin{aligned}\hat{y}_n(r) &= (L_n + rT_n)S_{n+r-12}(r = 1,2,\cdots,12)\\ &= (L_n + rT_n)S_{n+r-24}(r = 13,\cdots,24)\\ &= \cdots\end{aligned}$$

它与 Holt 法的差别在于用了一个适当的季节性因子来调整趋势预测值。在这里，调整是乘法的，但也可以是其他方式的调整。

水平（$L_n$）、趋势（$T_n$）和季节性分量（$S_n$）估计的更新方程是必需的。此外，这些方程也设定了“新的”和“旧的”加权平均。

$$L_n = \alpha_1 \frac{y_n}{S_{n-12}} + (1 - \alpha_1)(L_{n-1} + T_{n-1})$$
$$T_n = \alpha_2(L_n - L_{n-1}) + (1 - \alpha_2)T_{n-1}$$
$$S_n = \alpha_3 \frac{y_n}{L_n} + (1 - \alpha_3)S_{n-12}$$

三个平滑常数 $\alpha_1$，$\alpha_2$ 和 $\alpha_3$ 决定了观察是如何加权的。前两个方程与霍尔特的方程相似，唯一的差别在于最近的观察因季节性而作了调整（$y_n/S_{n-12}$）。最后一个方程中的季节性因子的更新是最后一个可用的“旧的”季节性因子（$S_{n-12}$）和“新的”季节性因子的信息，$y_n/L_n$ 的加权平均。

平滑常数经由将单时间区间的预测误差的平方和最小化求得。这一计算始于水平、趋势和季节性分量的初始值，估计值递归地更新，直至到达最后一个可用的时间周期为止，而预测值则是计算得到的。虽然计算更加复杂了，但在概念上却与简单的版本一般无异。

## 预测区间

点预测是非常重要的，但是如果它不附带一个置信【Confidence】水平则意义甚微。未来的观察 $y_{n+r}$ 的，一个 $r$ 步的向前预测的95%左右的置信区间由下式给出：

$$\hat{y}_n(r) \pm 2s_r$$

式中的均方根误差【Root Mean Square Error】

$$s_r = \sqrt{\sum_{t=r+1}^{n} \frac{[y_t - \hat{y}_{t-r}(r)]^2}{n-r}}$$

是从“观察到的”$r$ 个时间区间向前预测的误差 $y_{r+1}-\hat{y}_1(r)$，$y_{r+2}-\hat{y}_2(r)$，…，$y_n-\hat{y}_{n-r}(r)$ 计算得到的。

## 博克斯-詹金斯 ARIMA 模型

指数平滑法是一种在应用于一个时间系列时能得到预测值的程序。博克斯-詹金斯 **ARIMA**【Box-Jenkins Arima】模型是一种基于模型的方法。首先，我们要为潜在的产生系列的随机过程确定模型，然后我们再用模型导出最佳的预测值。

自回归求和移动平均（ARIMA 模型）把自相关观察 $y_t$ 看成在不相关的随机变量 $a_t$ 上的线性滤波的结果。将滤波写作多项式的比率后导致差分方程

$$\varphi(B)y_t = \phi(B)(1-B)^d y_t = \theta_0 + \theta(B)a_t$$

式中 $B$ 是二次变换运算符，$B^m y_t = y_{t-m}$；$a_t$ 是有常方差的不相关的、未观察到的误差；$\varphi(B)=\phi(B)(1-B)^d = 1-\varphi_1 B-\varphi_2 B^2-\cdots-\varphi_{p+d}B^{p+d}$ 是 $p+d$ 阶的广义自回归运算符；$d$ 是将非稳定系列转变为稳定系列所需要的自由度；$\phi(B)=1-\phi_1 B-\cdots-\phi_p B^p$ 是 $p$ 阶的稳定自回归运算符；$\theta(B)=1-\theta_1 B-\cdots-\theta_q B^q$ 是 $q$ 阶的可逆移动平均运算符；$\theta_0$ 是在 $d\geqslant 1$ 时引进确定性趋势的常数。

## 特例

一阶 ARIMA(1, 0, 0)模型，$(1-\phi B)y_t=\theta_0+a_t$ 或 $y_t=\theta_0+\phi y_{t-1}+a_t$ 是当前观察在其滞后上的回归。ARIMA($p$, 0, 0)模型是在时间 $t$ 的观察在 $p$ 个以前的滞后上的回归。ARIMA(0, 1, 0)的随机游走模型，$(1-B)y_t=\theta_0+a_t$ 或 $y_t=\theta_0+y_{t-1}+a_t$ 对于给随机的非稳定性建模很有用，常数 $\theta_0$ 引进了额外的确定性的漂移。ARIMA(0, 1, 1)模型，$(1-B)y_t=(1-\theta B)a_t$ 或 $y_t=y_{t-1}+a_t-\theta a_{t-1}$ 是另一种普遍使用的模型。扩展到包括季节性情形的模型也已经开发出来了。

博克斯和詹金斯(Box & Jenkins，1976)对这些模型的性质进行了研究，并拟就了有用的建模策略大纲。模型的阶($p$, $d$, $q$)是从系列的自相关【Autocorrelations】和偏自相关模式求得的，模型的参数是用最大似然法【Maximum Likelihood】估计的，而模型的精确度则是通过残差诊断评估的。

## ARIMA 模型隐含的预测

通过取未来观察的条件期望值求得未来观察的最佳(最小均方误差)预测值。这些条件期望值可从 ARIMA 模型的差分方程中导出。我们用二阶自回归模型来阐明这一点。

$$y_t = \theta_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + a_t$$

未来观察($y_{n+1}$，$y_{n+2}$，…)的条件期望值是

$$\hat{y}_n(1) = \theta_0 + \phi_1 y_n + \phi_2 y_{n-1}$$
$$\hat{y}_n(2) = \theta_0 + \phi_1 \hat{y}_n(1) + \phi_2 y_n$$
$$\hat{y}_n(r) = \theta_0 + \phi_1 \hat{y}_n(r-1) + \phi_2 \hat{y}_n(r-2) \quad (r \geqslant 3)$$

借助最终预测函数(eventual forecast function)我们得以更好地理解预测。

ARIMA($p$, $d$, $q$)模型的最终预测函数由下式给出

$$\hat{y}_n(r)=f_1(r)\beta_1^{(n)}+f_2(r)\beta_2^{(n)}+\cdots+f_{p+d}(r)\beta_{p+d}^{(n)}=f(r)^{\mathrm{tr}}\beta^{(n)}(r>[q-(p+d)])$$

式中的$f(r)=[f_1(r),f_2(r),\cdots,f_{p+d}(r)]^{\mathrm{tr}}$是预测函数的列向量,它只取决于广义自回归运算符(tr 表示向量转置)。例如,在 ARIMA(0, 2, $q$)模型中,预测函数由$f_1(r)=1$和$f_2(r)=r$给出,而最终预测函数则是线性的。

系数$\beta^{(n)}=[\beta_1^{(n)},\beta_2^{(n)},\cdots,\beta_{p+d}^{(n)}]^{\mathrm{tr}}$的向量是根据

$$\beta^{(n)}=L\beta^{(n-1)}+h[y_n-\hat{y}_{n-1}(1)]$$

用每一个新的观察进行更新的。转换矩阵$L$是预测函数$f(r)$的结果,而向量$h$则取决于 ARIMA 模型中的参数。

### ARIMA 预测法与指数平滑法之间的关系

ARIMA 模型隐含的最终预测函数和指数平滑法的函数存在着一定的相似之处。自回归分量和差分都隐含着 ARIMA 预测法的预测函数。在最终预测函数中和指数平滑法中的系数更新可以是相同的,如果我们选择了某些特定的 ARIMA 模型的话。

ARIMA 模型更具普遍性。约束 ARIMA 模型的预测与指数平滑的预测相符。ARIMA(0,1,1)模型的预测,$(1-B)y_t=(1-\theta B)a_t$与那些简单指数平滑的预测相符,条件是$\theta=1-\alpha$。ARIMA(0,2,2)模型,$(1-B)^2y_t=(1-\theta_1B-\theta_2B^2)a_t$与 Holt 指数平滑的预测相同,条件是$\theta_1=2-\alpha_1(1+\alpha_2)$和$\theta_2=-(1-\alpha_1)$。季节性预测方法之间的关系则更加复杂(Abraham & Ledolter, 1986)。

尽管从理论上讲,这些预测法各有长短,但是在许多实际的应用中,预测法之间的"较量"显示简单指数平滑法则更胜更为普遍的 ARIMA 模型一筹(Makrida-kis & Hibbon, 2000)。更为复杂的模型和所有可能的不正确模型都缺乏稳健性和用不必要的参数拟合"不吝啬的"模型带来的问题都能说明这一点。

## 本词条未曾涵盖的重要问题

许多其他的单变量预测方法已经被开发出来,但本词条却未能涵盖它们。在哈维(Harvey,1990)的著作中出现的结构时间系列模型和在财政金融的应用中发展起来的随机模型(参见 Tsay,2002)则是值得我们注意的。本词条主要介绍了数量方法,但未曾提及在只有很少或完全没有数据时是很重要的判断预测方法(参见 Armstrong, 2001)。

单变量时间系列只使用自身过去的历史。我们可以期望在将来自其他系列$X$的信息合并进来时,$Y$的预测将得到改进。如果$y_t$为当前或滞后的($x$值)影响,那么我们就可以使用转移函数模型,如果存在强的领先-滞后关系,那么预测便将因此而改进。$y_t$与$x_t$的同期联系本身却收益不多,因为为了预测$Y$我们需要$X$的一个预测值。多变量时间系列模型,特别是多元自回归模型是比较适宜的模型,如果系列之间的反馈存在的话。计量经济学【Econometric】模型使用许多方程系统来代表经济变量之间的关系。计量经济模型的目的在于提供条件或"假设分析"的预测值,并将之与单变量预测法提供的无条件预测作比较。

有许多不同的预测方法可资使用。每种方法各有长短,且其所依据的假定【Assumption】却常常被违反。理论和经验的证据表明,各种预测法的组合往往都具有一定的长处(*International Journal of Forecasting*, No.4, 1989)。

## 计算机软件

可资利用的性能优良的用于时间系列预测的软件有很多。所有主要的统计软件包(如SAS,SPSS和Minitab)都包含有用的时间系列预测模型。此外,若干种特定用途的程序(如ForecastPro和SCA)都包含有“自动的”程序,在计算过程中不需要用户做过多的调试,便可自行确定最适当的方法。这一点对大规模的预测应用是很重要的,因为在这样的预测中,往往涉及成千上万个系列的预报。

## 建议阅读的书目

由亚伯拉罕和莱道尔特(Abraham & Ledolter,1983),纽博尔德和博斯(Newbold & Bos,1994)及迪博尔德(Diebold,2001)撰写的著作,是这一领域的优秀的导论。由阿姆斯特朗(Armstrong,2001)编撰的《预测原理》(*Principles of Forecasting*)一书包括了普遍使用的预测方法的摘要和参考书目,内容涉及定量和定性(判断)两种预测程序。《预测期刊》(*Journal of Forecasting*)和《国际预测期刊》(*International Journal of Forecasting*)是两种有关预测问题的专业期刊。

——Johannes Ledolter

(沈崇麟译校)

## 参考文献


Abraham, B., & Ledolter, J. (1983). *Statistical methods for forecasting*. New York: Wiley.

Abraham, B., & Ledolter, J. (1986). Forecast functions implied by ARIMA models and other related forecast procedures. *International Statistical Review*, *54*, 51-66.

Armstrong, J. S. (Ed.). (2001). *Principles of forecasting*. Boston: Kluwer.

Box, G. E. P., & Jenkins, G. M. (1976). *Time series analysis: Forecasting and control* (rev. ed.). San Francisco: Holden-Day.

Diebold, F. X. (2001). *Elements of forecasting* (2nd ed.). Cincinnati, OH: South-Western.

Harvey, A. C. (1990). *Forecasting, structural time series models, and the Kalman filter*. New York: Cambridge University Press.

Makridakis, S., & Hibbon, M. (2000). The M3-competition: Results, conclusions, and implications. *International Journal of Forecasting*, *16*, 451-476.

Newbold, P., & Bos, T. (1994). *Introductory business forecasting* (2nd ed.). Cincinnati, OH: South-Western.

Tsay, R. S. (2002). *Analysis of financial time series*. New York: Wiley.


# 预想(Foreshadowing)

一些从事定量研究的学者用这个术语来指称人们在进入某个领域之前就指定其感兴趣的议题的过程。正如英国社会人类学家布罗尼斯拉夫·马林诺夫斯基所说:“先入之见在任何科学工作中都是有害的,但是预想问题是科学思想家的最大的天分所在,这些问题首先是观察者通过理论研究揭示出来的。”(Malinowski, 1922, pp. 8-9;引自Hammersley & Atkinson, 1995, pp. 24-25)因此,不同于定性研究者有时所主张的开放式方法,马林诺夫斯基及其追随者建议,去田野时要对需要调查的某些议题有明确意识。

——Alan Bryman

(杨可译 沈崇麟校)

## 参考文献

Hammersley, M., & Atkinson, P. (1995). *Ethnography: Principles in practice*. London: Routledge.

Malinowski, B. (1922). *Argonauts of the Western Pacific*. London: Routledge & Kegan Paul.

# 福柯式话语分析(Foucaldian Discourse Analysis)

米歇尔·福柯(Foucault, 1998/2002)在其著作《性史》(*The History of Sexuality*)与《知识考古学》(*The Archeology of Knowledge*)中发展了一种复杂、精致的话语理论。尽管他的作品囊括的经验主题范围很广,在方法上也有大量的变化,在其作品中有三个重复出现的主题,给他各种不同的兴趣提供了一定的连贯性:权力、知识和主体性。理解福柯式话语分析就从理解他对这三个概念的概念化【Conceptualization】及其关系入手。

福柯把话语或话语构成定义为构成其讲述物体的知识体系。换句话说,话语不只是描述社会世界,它通过给特定的社会现象分类、赋予意义的办法来让社会现象存在,否则它就是无意义的现实。任何时间、任何地点的话语构成的本质,就是在任何时间点上都存在的权力关系(以及反抗)、社会客体与认同,以及说和做的可能性的来源。

所有的话语都被定义成了一套规则或原则——“构成原则”——这导致特定客体的出现,它们构成了可识别的社会世界。话语制定了“可能性的条件”,这些条件决定了可以说什么、谁来说、什么时候说。任何陈述都不是偶然发生的,话语分析者的任务就是分析为什么出现的是某个具体陈述而不是另一个,它又是怎样出现的。

对福柯来说,话语——或者至少是它举例说明的知识——不能脱离开权力。权力嵌入在知识中,正如“权力/知识”这一对词简要概括的一样,任何知识体系都构成了一种权力的系统。在构建可能的身份认同、观念、社会客体时,权力的环境也形成了。反过来,权力则带来了新式的知识,生产了新的社会客体。

权力并非与某些代理连接的事物,而是代表着由话语所构成的知识体系决定的一个复杂的关系网络。因此,正是话语环境而非任何个体行动者的主体性在对政治策略的性质产生影响。事实上,对福柯而言,那种认为主体的有意识行动在某种程度上不是由话语所决定的观念恰恰说反了。

话语以这种方式决定了可能的条件,“规范”着主体,行动者因此在特定的话语环境中被理解并理解他们自己。由此,话语形塑了个人的主体性。福柯将主体去中心化了:不存在本质或真的主体,只有一个权力/知识体系,它构成了嵌入其中的个体的主体性。

福柯的工作对社会研究影响巨大,尤其是在对话语和权力之关系的理解方面。同时,他的经验方法具有一种极为特殊的风格,在经验层面上除了他自己的研究,还没有发现太多的后继者。无论如何,他的工作影响极为深远,假以时日,必定会对社会科学产生更为重要的贡献。

——Nelson Phillips

(杨可译　沈崇麟校)

## 参考文献

Foucault, M. (1998). *The history of sexuality: The will to knowledge* (Vol. 1). London: Penguin.

Foucault, M. (2002). *Archeology of knowledge*. London: Routledge.

## 研究中的欺骗(Fraud in Research)

自然科学(尤其是生物学和医学)中的欺骗并不少见(Broad & Wade, 1982)。而社会科学研究中的欺骗据说是相对较少的。当然,这种说法取决于我们如何定义“欺骗”。无论在普遍意义上,还是在特定案例中,这都是一个充满争议的概念。这种固有的不确定性有两个来源:一个是在这一概念下可能出现的活动的转移和模糊性,另一个则是“成功的悖论”。

研究欺骗的类型学分类是非常丰富的。1830年,查尔斯·巴比奇对于“篡改”“杜撰”和“伪造”的违法所作的描述(Broad & Wade, 1982, pp. 29-30)是早期经常被引用的一种“行动型”方向的欺骗案例;另外一种被称为“目的型”方向,例如,将欺骗与恶作剧区分开来,欺骗的实施带有严肃和固定的意图,而恶作剧往往带有玩笑的意味,并且是一种临时行为。因此,欺骗的实施者比恶作剧的实施者更加恶名昭著。

然而,故意欺骗本身就是一种矛盾体。由于任何欺骗的实例都必须经过发现才能被认定,而发现本身意味着欺骗的失败,因此成功的欺骗是无法被感知存在的——或者说是不存在的。在某些评论家,诸如西里尔·伯特晚期的一位追随者罗伯特·乔因森(见下)看来,这种观点是有问题的:“‘一个成功的信任骗子是恰到好处地使人信服他的诚实。’如果我们接受这种奇怪的论调,那么任何赢得信任的人都应该称得上是一个成功的诚信的骗子。”(Joynson, 1989, p. 302)严格来讲,这种“成功的悖论”所指的是,那些诚实的工作和活动是无法与那些不诚实但却没有被发现的行为区分开的。这意味着如果我们想要了解欺骗现象,非常重要的一点是,我们要了解欺骗的被发现工作(揭露真相)(Ashmore, 1995)。

为什么在社会科学研究纯粹的欺骗较为罕见呢?一个原因是发现欺骗的机会是比较少的,且这样的工作都要耗费很多资源。尽管我们可以明显地观察到生物学家约翰·达斯“1981年5月的一个晚上公然[伪造]了原始资料”(Broad & Wade, 1982, p.14),但却没有诸如这样的简单揭露欺骗的技术,可为玛格丽特·米德(Special section, 1983)这样的人类学家,甚至西里尔·伯特这样的定量心理学家定罪。尽管伯特确实被利昂·卡明等人指责他无耻地伪造原始资料,但这种揭露真相的努力是不稳定的——因为之后可能会有其他人通过“元揭露”工作,来指责这些揭露者也犯了同样的罪行,在面对这样的指责时,之前的揭露者也是十分脆弱的(Joynson, 1989; Mackintosh, 1995)。社会科学的风格越定性和发散,欺骗被发现的可能性就越小,除非是学术剽窃。

——Malcolm Ashmore

(郑晓娟译　沈崇麟校)

### 参考文献


Ashmore, M. (1995). Fraud by numbers: Quantitative rhetoric in the Piltdown forgery discovery. *South Atlantic Quarterly*, *94* (2), 591-618.

Broad, W., & Wade, N. (1982). *Betrayers of the truth: Fraud and deceit in the halls of science*. New York: Simon & Schuster.

Joynson, R. B. (1989). *The Burt affair*. London: Routledge.

Mackintosh, N. J. (1995). *Cyril Burt: Fraud or framed*? Oxford, UK: Oxford University Press.

Special section: Speaking in the name of the real: Freeman and Mead on Samoa. (1983). *American Anthropologist*, *85*, 908-947.

## 自由联想访谈(Free Association Interviewing)

自由联想访谈(FANI)是一种将叙述重点和自由联想的心理分析原则结合在一起的,生产和分析访谈资料的定性研究方法。这些引起叙述的问题会鼓励被访者回忆那些人生经历中的特定事件。与广义的记述不同,这些访谈具有情感的共鸣。自由联想原则基于这种理念:一些看法之间无意中激发的联系,比这些看法的内容本身更能体现出被访者述说中的情感意义(详见心理分析法【Psychoanalytic Method】)。因此,它尤其适合用来探索那些能够体现被访者认同或触及被访者情绪和情感的问题(与之形成对比的是那些经过设计的用以引出观点或事实的方法)。

大部分社会科学经验研究方法都假定,研究主体都对自己有足够的了解,因此能够给出可靠的记述。相反,FANI 则认为,对焦虑的防御是被研究主体甚至也是所有主体(包括研究者)的核心特征。由于焦虑是研究关系的主要特征,并且或多或少取决于话题、情境和关系的融洽程度,对焦虑的防御会潜在地危及被访者了解自身的行动、目的和关系的能力。这种无意识的防御是一个拥有心理社会属性的主体的特征之一,是由防御、欲望和精神冲突所形塑的独特生平经历所产生的精神维度的一部分。这些无意识防御也与由共享的社会世界所构成的社会维度相互影响,这里所说的共享社会世界是主体之间的过程、真实事件和赋予事件意义的话语这三者的结合。

这个研究方法是在一项研究犯罪过程中焦虑和恐惧之间关系的项目中,通过对传记叙事解释法进行改进而发展起来的(详见传记叙事解释法【Biographic Narrative Interpretive Method】)(Hollway & Jefferson, 2000)。研究者采用了对特定事件的叙述问题格式("你可以告诉我当你……的时刻吗?)来阻止那些被半结构化访谈【Semistructured Interview】频繁引出的一般化和理性化的回答。研究者使用开放式【Open-Ended】问题,避免"为什么"类型的问题。被访者的回答能够一直沿着自己的讲述顺序、自己的叙述方式推进。这些访谈规则的设计使得研究参与者能够一直置身于自己的意义框架体系,而不是进入访谈者的意义体系中。

使用格式塔原则的 FANI,需要在解释某一特定片段的意义时,将收集的整体资料置于脑中。与心理分析的准则相一致,研究者对文本中的不连贯和分歧对立,例如,情绪基调的变化、矛盾之处和刻意回避更加保持警觉。

在访谈和资料分析过程中,访谈者将自己的感受当作资料,通过移情和反移情的心理分析原则来确认自己的情感投入。每位被访者都要接受两次访谈。两位研究者会在开始第二次访谈前听第一次访谈的录音磁带,这样一个注意留心那些联系、关联以及矛盾之处的意义的过程,将有助于构想第二次访谈的问题。

——Wendy Hollway
Tony Jefferson
(郑晓娟译　沈崇麟校)

### 参考文献

Hollway, W., & Jefferson, T. (2000). *Doing qualitative research differently: Free association, narrative and the interview method*. London: Sage.

## 频数分布(Frequency Distribution)

频数分布显示观察落入类别的每一类或一个变量【Variable】的各个值域的数目。频数分布常常用直方图【Histogram】、频数分布多边图【Frequency Polygon】,或频数表来描绘。在这样的分布用比例或百分比来表述时,它便称为相对频数分布【Relative Frequency Distribution】。在分布代表增加的频数时,它便称为累积频数分布。

——Tim Futing Liao

(沈崇麟译校)

## 频数分布多边图(Frequency Polygon)

频数分布多边图是若干种用来描述一组比较大的定量数据图形中的一种。图1是一张使用一组成年男性身高绘制的频数分布多边图的实例。

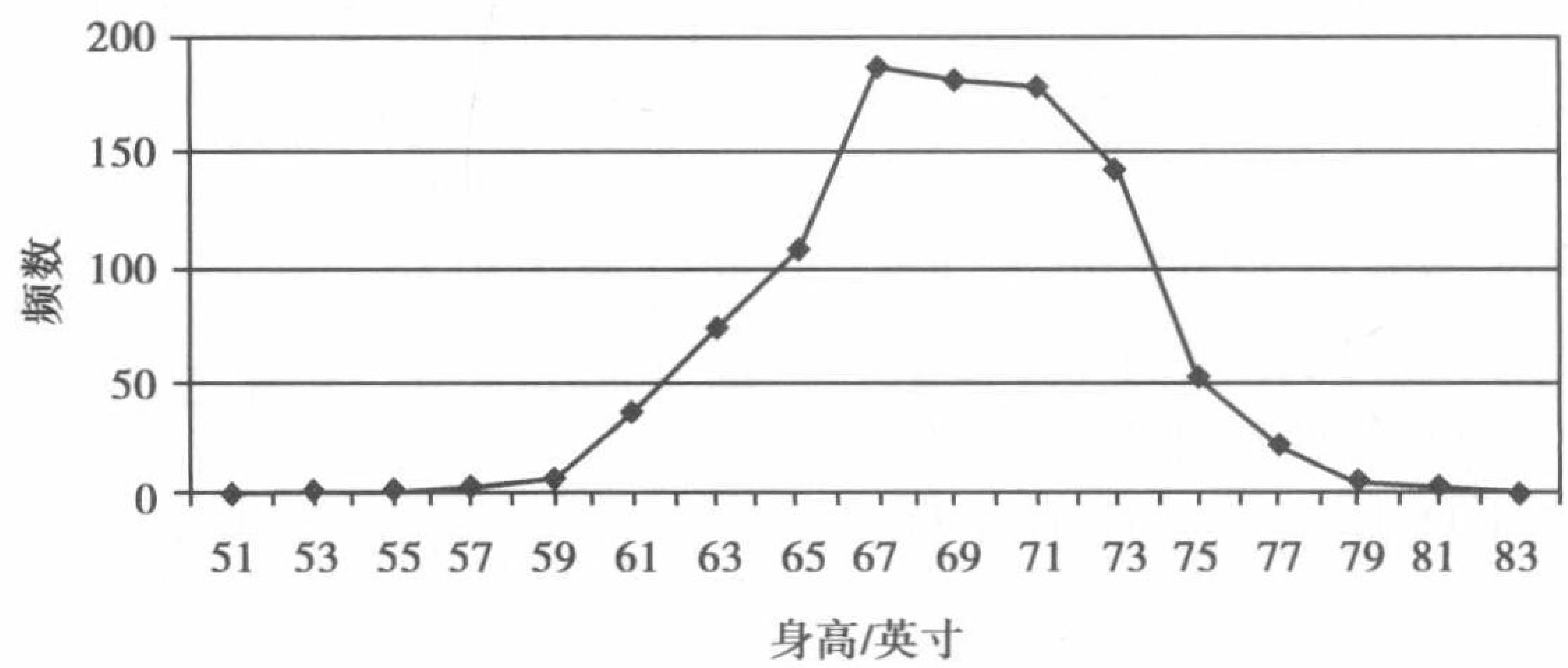



图1　区间为两英寸的,身高在52~82英寸的成年男性身高频数分布图

为了绘制一张频数分布多边图,我们首先要将数据由高到低进行排序,然后再将它们分组置于连续的区间。区间置于图形的 $x$ 轴上,由它们的中点标示。$y$ 轴则是一个频数量度,它显示落入每一个区间的数据点的数目。该尺度的范围从最小的0到一个超过模式区间的最大值。在数字域内用直线连接标志每一区间的频数的点。这些线条应当沿 $x$ 轴延伸,这意味着在数据集的低端和高端之间的空的区间也应该包括在内。

在数据的区间不等于数据集中不同的值的数目时,频数分布多边图被用来表示定量的数据。对于连续数据而言,情况总是这样。在数据是离散的时候,则有可能在 $x$ 轴上表示数据集中的每一个值。在这样的情形中,则可使用直方图【Histogram】,而不是频数分布多边图。二者的重要区别在于,在频数分布多边图中,用直线将点连接起来,允许 $x$ 轴的值的频数插入在非中点处。在数据没有在 $x$ 轴上标示的值之外的其他值时,只需用条(列)标出即可。

如果 $y$ 轴用来表示比例而非频数,那么图形就称为相对频数分布多边图。对某些数据而言,显示数据区间的累积频数比绝对频数更有用。在这种情况下,我们可用累积频数分布多边图(或肩形图)来描述数据集。

——Kenneth O. McGraw

(沈崇麟译校)

# Friedman 检验(Friedman Test)

弗里德曼(Friedman,1937)检验提供了一种双向设计等治疗【Treatment】效应的虚无假设【Null Hypothesis】检验。这种实验设计【Experimental Design】也称为平衡完全区组设计【Block Design】、随机化完全区组设计和重复测量设计【Repeated-Measures Design】。

假设我们有 $t$ 种治疗、$b$ 个区组和 $N$ 个实验单位,在这里,$N$ 是乘积 $bt$ 的倍数。为了方便起见,我们假定 $N = bt$,这样在每个区组就会有恰好 $t$ 个实验单位。$t$ 个治疗被随机分配【Randomly Assigned】给第 $i$ 个区组内的 $t$ 个单位,得到观察 $X_{i1}, X_{i2}, \cdots, X_{it}$。为了进行这一检验,我们将在第 $i$ 个区组的这些观察用 1, 2,⋯, $t$ 这样的整数序号,从最小到最大排列起来。对每一个区组中的观察都重复进行同样的排序。令 $R_{ij}$ 为分配给区组 $i$ 中的实验单位受到的治疗 $j$ 的等级。完全的等级数据组可用下面的双向表来表示:

治疗

| 区组 | 1 | 2 | ⋯ | $t$ | 总和 |
|---|---|---|---|---|---|
| 1 | $R_{11}$ | $R_{12}$ | ⋯ | $R_{1t}$ | $t(t+1)/2$ |
| 2 | $R_{21}$ | $R_{22}$ | ⋯ | $R_{2t}$ | $t(t+1)/2$ |
| ⋮ | ⋮ | ⋮ | | ⋮ | ⋮ |
| $b$ | $R_{b1}$ | $R_{b2}$ | ⋯ | $R_{bt}$ | $t(t+1)/2$ |
| 总和 | $R_1$ | $R_2$ | ⋯ | $R_t$ | $N$ |

我们注意到,每个行和都是 $1+2+\cdots+t=t(t+1)/2$。如果治疗效应确实都相同,就像在虚无假设下一样,那么每一行的秩无非就是 1,2,⋯,$t$ 的随机排列,而列和将趋于相等。但是列和的合计必定等于 $N=bt(t+1)/2$,因而每个列和必定等于 $bt(t+1)/2$ 除以 $t$,或 $b(t+1)/2$。检验统计量 $S$ 是实际的列和与虚无假设下的期望和的偏差的平方,或

$$S = \sum_{i=1}^{t}\left[R_i - \frac{b(t+1)}{2}\right]^2 = \sum_{i=1}^{t} R_i^2 - \frac{b^2t(t+1)^2}{4}$$

如果在治疗效应之间有一个确定的次序,某一列的所有的秩将是数 $1b$, $2b$,⋯,$tb$ 的某个排列。它可能等于 1,在某个其他的列中的所有的秩将可能会是 2,如此等等,而列和则会显示,这些列和将会产生 $S$ 的最大的可能值。因此,等治疗效应的虚无假设,便会因大的 $S$ 值而被拒绝,而治疗之间有一种确定的次序的备择假设则得到了支持。

小的 $b$ 和 $t$ 值的精确 $S$ 虚无分布已制成表格,并在吉布森(Gibbons,1993, 1997)的著作中重制。对于大样本,检验统计量是 $Q=12S/bt(t+1)$,它近似于有 $t-1$ 个自由度【Degrees of Freedom】的卡方分布。

举一个数字的例子,假设我们想比较外币兑换美元的中位数的日波动程度。这些外币是澳大利亚元(AD)、英镑(BP)、日元(JY)和欧元(EE)。波动程度用一周中的每一天计算的偏斜系数来量度。表 1 列出了相对值,表中的秩 1=最小值。我们有 $t=5$ 种治疗(一周中的天数)和 $b=4$ 个区组(货币)。我们算得 $S=138$ 和 $Q=13.8$,有四个自由度。

近似的 $p$ 值是 $p<0.01$,因而拒绝等中位数波动的虚无假设。它似乎显示出一种显著的日效应,因为最大的波动发生在周的中间,而最小的波动则发生在周初和周末。

表 1 货币波动数据

| 货币 | 星期一 | 星期二 | 星期三 | 星期四 | 星期五 |
|---|---|---|---|---|---|
| AD | 2 | 3 | 5 | 4 | 1 |
| BP | 2 | 3 | 4 | 5 | 1 |
| JY | 1 | 4 | 3 | 5 | 2 |
| EE | 1 | 3 | 4 | 5 | 2 |
| 总计 | 6 | 13 | 16 | 19 | 6 |

——Jean D. Gibbons

(沈崇麟译校)

## 参考文献

Friedman, M. (1937). The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *Journal of the American Statistical Association*, *32*, 675-701.

Gibbons, J. D. (1993). *Nonparametric statistics: An introduction.* Newbury Park, CA: Sage.

Gibbons, J. D. (1997). *Nonparametric methods for quantitative analysis* (3rd ed.). Columbus, OH: American Sciences Press.

# 函数(Function)

一个特定的现象或过程的运作方式的科学解释,一般都始于基本的行为主体和涉及的基本量的识别——也即变量和变量所联系的实体的识别。我们立即就要面对的挑战是这些基本量之间的关系。许多这样的关系都被概括和体现在下式这样的数学函数中,即

$$y = f(x_1, \cdots, x_n)$$

式中的 $y$ 和各个 $x$ 是基本的量,现在则被分别称为因变量【Dependent Variable】和自变量【Independent Variable】。

当我们开始思考这些量之间的关系时,其步骤包括:考虑哪些变量是作为哪些产出投入的,并将注意力限制在一个函数;考虑产出是否随每一个投入的增加而增加或下降;增长率是不变、上升还是下降的。当然,我们也要思考函数的域和范围以及那些与基本数量的数学表示紧密相连的元素。不仅如此,有时投入和产出之间的关系不是单调的,因为效应在某个点改变了方向。

已经证明,对一个特定的过程的运作方式的理解是逐渐加深的——经历从没有函数到广义函数,再到特定函数这样一个过程。随着这一问题变得很清楚,我们便会认识到某些变量应归入方程(作为各个 $x$),而另外一些变量则不应归入方程;$y$ 随着某些 $x$ 上升,随着其他的 $x$ 下降,而与另一些 $x$ 则不是单调关联;涉及不同的 $x$,变化率也有所不同。一般来讲,理论和经验的文献,即使在前函数阶段也对这些特性已经有了深入的见解。一种设定函数的路数是在文献中确定一个函数代表一种特定的关系所应当满足的那些条件。即使在一个特定的函数经由直觉和经验工作引起我们的注意时,设法论证该函数的特性与文献中重要内容之间的联系,以夯实

它的基础也是很重要的。

## 例子

分配-报应公正研究确定了两种行动者：一种是观察者和被报答者（他可以是，但不一定是观察者）和三个量；另一种是被报答者的实际报偿，观察者对被报答者公正报答的理念和观察者对于被报答者被公正或不公正地报答的判断，如果报偿不公正，那么报偿过高或过低的程度是多少？通常，我们都将这种评估称为公正评估。文献给我们提供了很多设定公正评估函数的颇有见地的观点。第一，文献使我们清楚（对于商品）公正评估随着实际报偿的上升而上升，随着公正报偿的上升而下降。第二，文献用真数来表示公正评估，零表示十分公正点，负数和正数分别表示过低和过高报偿。第三，文献也提出了对公正评估函数合适的两个性质：（a）标度不变性，这使得公正评估独立于报偿的量度单位。（b）可加性，这意味着实际报偿在报偿评估上的效应独立于公正报偿的水平，反之亦然。前两个见解导致广义函数，而第三点则证明在基本报偿中，一个函数式唯一地既满足了标度不变性，也满足了可加性条件（Jasso, 1990）。

为了推论一种关系而设定函数的学科会有高回报。即使在它无法导出唯一的函数式时，它也能使我们得以科学地理解过程的运作方式。

在社会科学中，函数有着广泛的应用，这包括心理学中的心理函数，经济学中的功用、生产和消费函数，经济学和社会学中的收入函数，社会学和社会心理学中的认知和地位函数。随着知识的增长，越来越多的关系被设定为函数。霍曼斯（Homans，1967, p. 18）对设定函数的生动挑战是永远及时的。

函数可以作为理论和理论模型的起始假定【Assumption】，它也可以直接用于经验工作（Jasso, in press）。因此，函数把相干性赋予了一些局部域。

## 延伸阅读

定期精读数学手册（如 Bronshtein & Semendyaev, 1979/1985）的函数和微积分教科书（如 Courant, 1934/1988），帮助我们的思想作好面对那些模糊关系的准备（且能立即给我们以愉悦）。阅读社会科学中著名的函数及其发展，如费希纳（Fechner, 1860, 1877）关于知觉函数的著作和明瑟（Mincer, 1958, 1974）关于收入函数的著作都能起到同样的功效。

——Guillermina Jasso

（沈崇麟译校）

## 参考文献


Bronshtein, I. N., & Semendyaev, K. A. (1985). *Handbook of mathematics* (K. A. Hirsch, Trans. & Ed.). New York: Van Nostrand Reinhold. (Originally published in 1979.)

Courant, R. (1988). *Differential and integral calculus* (2 vols.) (E. J. McShane, Trans.). New York: Wiley. (Originally published in 1934.)

Fechner, G. T. (1860). *Elemente der psychophysik.* Leipzig: Breitkopf & Härtel.

Fechner, G. T. (1877). *In sachen der psychophysik.* Leipzig: Breitkopf & Härtel.

Homans, G. C. (1967). *The nature of social science.* New York: Harcourt, Brace, & World.

Jasso, G. (1990). Methods for the theoretical and empirical analysis of comparison processes. *Sociological Methodology*, *20*, 369-419.

Jasso, G. (in press). The tripartite structure of social science analysis. *Sociological Theory.*

Mincer, J. (1958). Investment in human capital and personal income distribution. *Journal of Political Economy*, *66*, 281-302.

Mincer, J. (1974). *Schooling, experience and earnings.* New York: Columbia University Press.

# 模糊集合理论(Fuzzy Set Theory)

模糊集合理论用于处理那些边界不清楚或“模糊的”集或类别。例如,一个物体不一定只是红的或不是红的,它可能是“淡红的”甚至是“黄绿的”(带一点红的绿色)。同样,一个政治科学家可能将一个政府定为与另一个政府差不多民主。模糊集合理论允许红色物体或民主政府集中的成员归属问题是一个程度问题。它给我们提供了一种处理同时有类别属性和维度属性的概念的分析框架。下面我们将向大家介绍模糊集合理论和模糊逻辑背后的基本理念。在介绍时只要可能,我们便会使用来自社会科学的例子进行阐述。

## 模糊集合的概念

一个集合中的归属成员可以用一个[0,1]区间中的值来表示。一个经典的清晰集合只允许有0(非正式成员)和1(正式成员)两个值。将一个男性的集合作为清晰集合处理,就必须在男性这一标准得到满足时,分配给每个人1,否则则分配0。另一方面,一个模糊集合则允许有0和1之间的值。因此,一个21岁的人可能作为一个“年轻成年”集合中的正式成员,但一个28岁的人则可以给他一个隶属度,位于0和1之间的某个地方。按照惯例,1/2表示一个既不是完全在集合内,也没有在集合外的元素。一个归属函数(membership functio)将一个或多个支集变量的值或陈述在[0,1]区间中的位置一一标明。图1显示了两个假设的归属函数,用年龄作为支集:“青年”和“青少年”。尽管青少年一般都比青年年轻,但是这两个集合是重叠的,且每个集合都允许有非零的归属度。

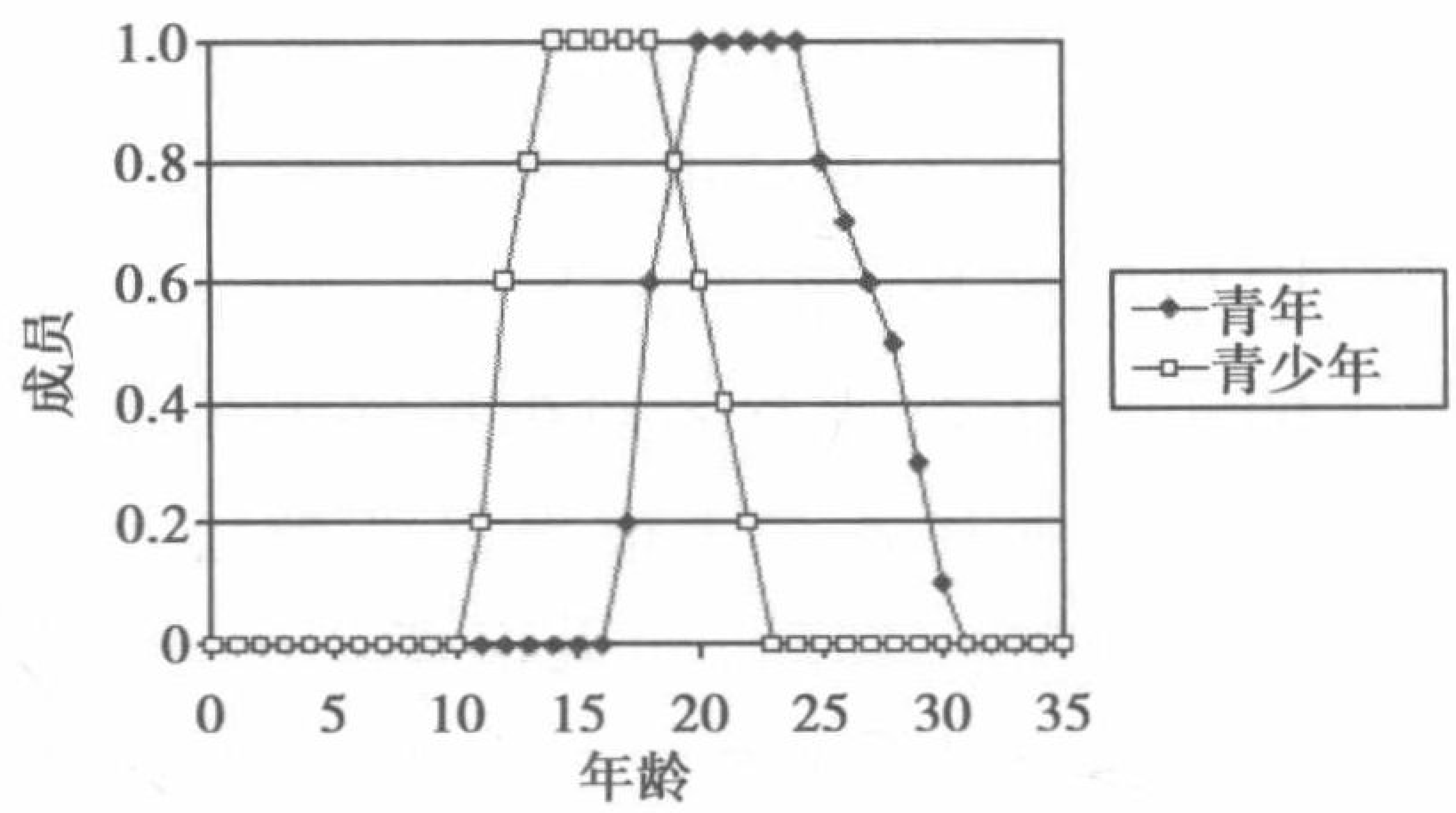



图1 “青年”和“青少年”的归属函数

虽然不存在为大家普遍接受的,有关模糊集合中的归属度的解释,但是却存在这几点共识。首先,归属度并非概率,它确实可以被赋予完全的确定性。不仅如此,与概率不同,跨一个完备的备择集合的归属度之和不必等于1。在某些场合,成员是用对某种原型的相似度来定义的,而在另一些场合,它可能意味着有关某个概念的兼容性或可能性。后者的一个例子是“几个”一词,被认为是一个模糊的语言数字。整数6被认为与“几个”完全相容,而3则被认为较少相容,因而它将被分配一个较低的归属值。

可能性这一概念也被用来描述归属度的特性。可能性提供了概率的上包封。例如,在一个社区中有45%的人有自行车,那么在该社区的一个随机调查中,在一个给定的日子发现一个用自行车的人的最大概率就是0.45,当然,实际的概率可能大大低于它。概率理论是一种以模糊集合理论为基础的框架,且与不确切的概率紧密相连(例如,低与高概率和信任函数)。

模糊性问题一直存在某些争论,人们普遍认为模糊性不是模棱两可、模糊不清、非特异性、不可靠或不一致。相反,模糊性涉及一种精确的等级性。例如,一个艺术家对于色彩的判断,都不是模棱两可的、非特异性的或带有贬义的模糊不清。人们提出了

各种模糊性量度,从概括性的信息论的熵到相对变异。所有这些量度的共同之处是这样一个概念:在它的更多成分的归属值接近0或1时,模糊就比较小;而在有更多的归属值接近1/2时,模糊则较大。

归属函数水平的量度也一直存在争论,尽管模糊集理论的一些重要方面甚至已被用于定序水平的归属函数。在许多应用中,我们可以对端点(0和1)作出合理的定义。但是却很难定义等间距中间归属值。一种有用的代表定序水平成员的方法是通过水平集合,它们是在归属尺度上的定序割点形成的嵌套的清晰集合。这些集合中的六种都由图1中的水平的网格线表示。例如在0.6的归属水平,青年集合包括18~27岁,而在归属水平1,青年则缩小到20~24岁。水平集合与(以上)累积分布函数密切相关。

最后,归属函数可能以其他集合中的状态或隶属度为条件。第一种情况发生在我们步行和驾车旅行时,模糊的距离“附近”便是一例。而将“高”用于描述男人和女人的身高,便是第二种情况的一个例子。

## 模糊集合的基本运算

令 $m_A(x)$ 为归属元素 $x$ 在集合 $A$ 中具有的归属的度。在不存在含糊不清时,下标或 $(x)$ 可以去掉。令 $\sim A$ 为 $A$ 的补集。模糊的否定式与概率【Probability】相同:

$$m_{\sim A} = 1 - m_A$$

根据图1中的归属函数,在青少年集合中的20岁的归属度是0.6,这样在补集中的归属度就是1-0.6=0.4。在 $A$ 和 $B$ 的交集中的归属度则由下式定义:

$$m_{A \cap B} = \min(m_A, m_B)$$

而 $A$ 和 $B$ 的合集的归属度则由下式定义:

$$m_{A \cup B} = \max(m_A, m_B)$$

在图1中,一个20岁的人既是青少年也是青年,归属度为 min(0.6,1)= 0.6。模糊集合的合集并和交集是非互补的。与加权线性组合不同,例如,如果 $m_A > m_B$,那么在 $m_A$ 中的一个增量,将不会补偿 $m_B$ 在 $m_{A \cup B}$ 中的归属度的下降的影响。在这一定义中,关键的假定的含义是集合 $A$ 和集合 $B$ 的函数有着完全的联合次序,这样比较运算的最小运算和最大运算都是允许的。

如果对于所有的 $x$ 而言,$m_A(x) > m_B(x)$,我们就说集合 $B$ 是集合 $A$ 的子集(集合 $A$ 包括集合 $B$)。注意这一定义也要求联合次序假定。模糊集合包含是清晰集合包含在概念上的一个推广,因此,直接与Guttman和Rasch量表)相关。尽管几乎不存在这种不平等性完全被满足这样的情况,但它在相当大的程度上得到满足这样的真实的例子是不难找到的。

图2显示了一个这样的例子,该图中共有83位回答人,它们标示了自己对命题“澳大利亚应当允许移民进入自己的国家”和“应当允许船民进入澳大利亚,并处理他们的申请”的同意程度。散点图显示,这种包含关系只有三个例外,位于图中的右上角。同意第二个陈述的人的集合,几乎都是那些同意第一个陈述的一个子集。

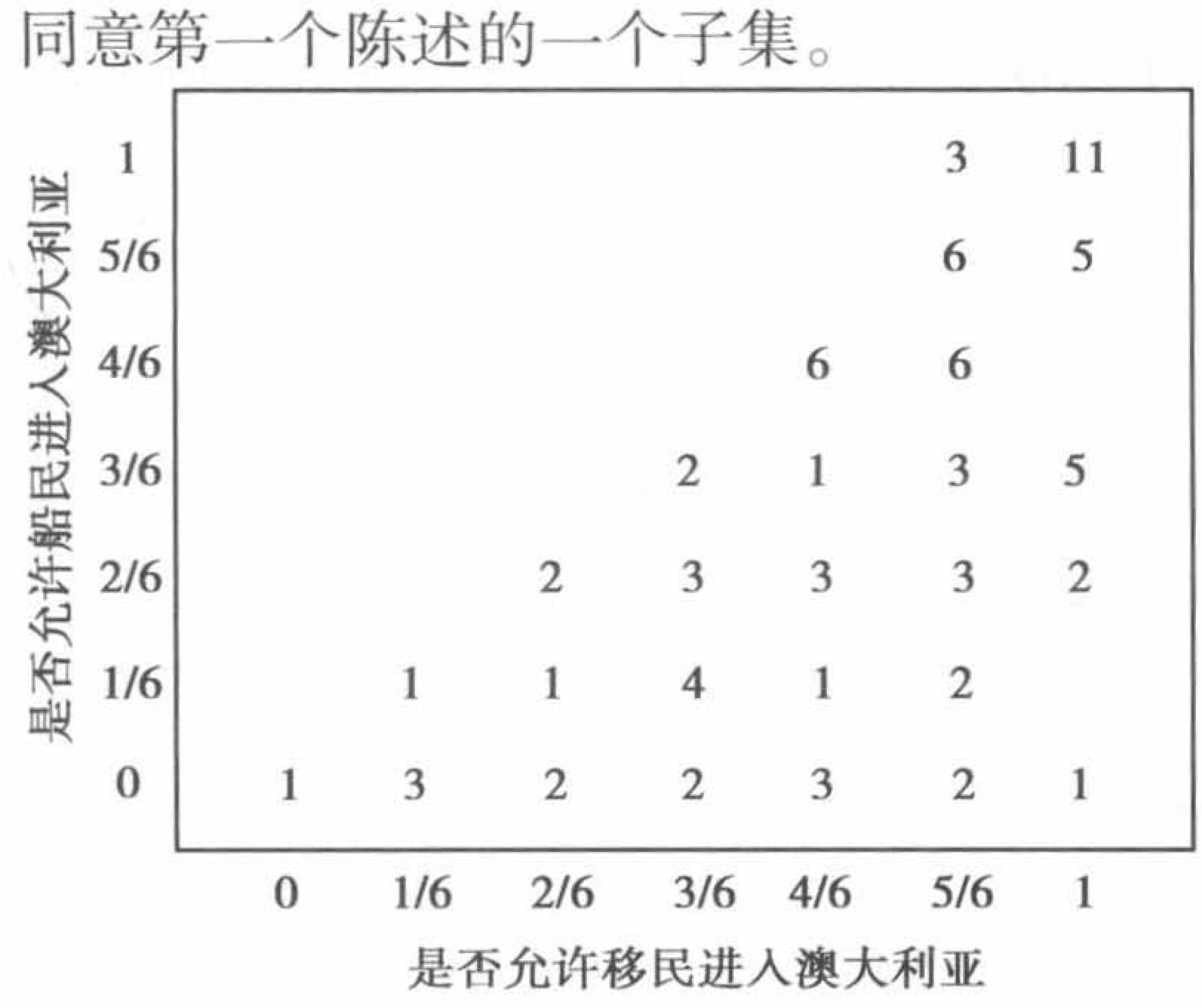



图2 模糊集合包含的例子

包含和模糊的概念被用来按照近似自然语言的限制语，如“非常”来解释归属函数的各种转换。一个转换如果降低了归属值，它便会使一个模糊集合集中，反之，则会使集合扩大。“非常高”的男人的集合可以说是“高的”男人集合的集中，而原来建议的集中运算符却是$[m_A(x)]^2$。虽然诸如这样的运算符一直在受到强烈的批评，被认为是不精确的口头限制语形式，但是它们却提供了某些转换的解释。

## 模糊逻辑

经典的布尔逻辑是双值的，每一命题非真即伪。多值逻辑至少允许有第三种陈述：“不确定的”或“可能的”。在布尔逻辑中，真值被限制为0(假)和1(真)，而三值逻辑将真值1/2分配给了第三种陈述。模糊逻辑对这种逻辑作了概括，方法是允许在[0, 1]的区间内有任何真值。

模糊逻辑原来的版本使用最小和最大运算符来计算逻辑合取和析取。用$t_P$和$t_Q$表示命题P和Q的真值，通常的逻辑基元则定义如下：

$$\sim P = 1 - t_P,$$
$$P \wedge Q = \min(t_P, t_Q),$$
$$P \vee Q = \max(t_P, t_Q),$$
$$P \Rightarrow Q = \sim P \vee (P \wedge Q)\ (\text{最大法则})\ \text{或}$$
$$= \sim P \vee Q\ (\text{运算法则}),\ \text{和}$$
$$P \Leftrightarrow Q = (P \Rightarrow Q) \wedge (Q \Rightarrow P)$$

只有在$t_P < t_Q$时，最大最小化和四则运算法则才会产生P⇒Q的不同真值，具体则如表1所示。

其结果是可能影响结果的四则运算法则的真值大于或等于最大最小化法则的真值。在更为普遍的备择模糊逻辑中，武卡谢维奇($L_1$)的版本会产生最宽大的可能影响结果的真值。在$L_1$中，最小和最大运算符被用于P∧Q和P∨Q，但是$P \Rightarrow Q = \min(1, 1-t_P+t_Q)$。因此，在$L_1$中，每当$t_P<t_Q$时，P⇒Q的真值为1，这与前述章节中的包含度的定义一致。

**表1　最大最小化和四则运算法则真值表**

| | | 最大最小化 | 四则运算 |
|---|---|---|---|
| $t_P$ | $t_Q$ | P⇒Q | P⇒Q |
| 0 | 0 | 1 | 1 |
| 0 | 1 | 1 | 1 |
| 1 | 0 | 0 | 0 |
| 1 | 1 | 1 | 1 |
| $p>$ | $q$ | $\max(1-p,q)$ | $\max(1-p,q)$ |
| $p<$ | $q$ | $\max(1-p,q)$ | $\max(1-p,q)$ |

举例说明，让我们回到前面图2所阐述的那个例子，在此例中，P=“是否允许船民进入澳大利亚”和Q=“是否允许移民进入澳大利亚”，要考虑的问题是P⇒Q。图3显示，用最大最小化和四则运算法则为图2的散点图中适当的命题产生的真值。它证明，如果将图2中83个个案的真值加在一起，最大最小化法则给出了总和68.33，而四则运算法则给出的是75.00。$L_1$将真值1分配给所有个案，除了那三个违反了$t_P<t_Q$的个案，而之所以给它们分配真值5/6，是因为总和为82.50。

模糊逻辑运算可以组合起来生成复合的可能影响结果的陈述，接下来它又可以形成非线性模型【Nonlinear Model】的基础。让我们来考虑下述三个命题：

I=给他人提供信息是重要的。

T=保持对密友的信任是重要的。

C=我确信我作出了正确的决定(公开或隐瞒消息)。

在一个职业调查中，我们对229个回答

最大最小化

| 是否允许船民进入澳大利亚 \ 是否允许移民进入澳大利亚 | 0 | 1/6 | 2/6 | 3/6 | 4/6 | 5/6 | 1 |
|---|---|---|---|---|---|---|---|
| 1 | | | | | | 5/6 | 1 |
| 5/6 | | | | | | 5/6 | 5/6 |
| 4/6 | | | | | 4/6 | 4/6 | 4/6 |
| 3/6 | | | | 3/6 | 3/6 | 3/6 | 3/6 |
| 2/6 | | | 4/6 | 4/6 | 4/6 | 4/6 | 4/6 |
| 1/6 | | 5/6 | 5/6 | 5/6 | 5/6 | 5/6 | 5/6 |
| 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

四则运算法则

| 是否允许船民进入澳大利亚 \ 是否允许移民进入澳大利亚 | 0 | 1/6 | 2/6 | 3/6 | 4/6 | 5/6 | 1 |
|---|---|---|---|---|---|---|---|
| 1 | | | | | | 5/6 | 1 |
| 5/6 | | | | | | 5/6 | 1 |
| 4/6 | | | | | 4/6 | 5/6 | 1 |
| 3/6 | | | | 3/6 | 4/6 | 5/6 | 1 |
| 2/6 | | | 4/6 | 4/6 | 4/6 | 5/6 | 1 |
| 1/6 | | 5/6 | 5/6 | 5/6 | 5/6 | 5/6 | 1 |
| 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

**图 3　最大最小化和四则运算法则真值表**

人就如何对待同事给他们透露的信息提出了三个问题。要求他们就这些问题中的每一个在一个 7 分量表上标出自己的同意程度，而我们在这里将把这些量表作为值的范围在 0~1 的真值处理。我们将要考查模糊逻辑模型。

$$C \Rightarrow I \vee T$$

这一模型提议，如果决策者是自信的，那么他们将会对问题 I 或 T（或者二者）同时打高分。这是一个非线性的模型，因为最大运算符是无法线性化的，而实际上，甚至是无法平滑或可微的。

表 2 中的前两张表显示了 I 与 C 和 T 与 C 的二元分布。第三张子表显示了 I∨T 和 C。目测告诉我们，那些高信任度的个案都被拉入了高 I∨T 区域，而那些低信任度的个案则沿 I∨T 范围非常平坦地散布。

**表 2　充分性和联合必要性**

I＝给他人提供的信息是重要的

| | 0 | 1/6 | 2/6 | 3/6 | 4/6 | 5/6 | 1 | 总计 |
|---|---|---|---|---|---|---|---|---|
| 1 | 8 | 6 | 4 | 4 | 3 | 7 | 36 | 68 |
| 5/6 | | 4 | 12 | 7 | 8 | 28 | 9 | 68 |
| 4/6 | | 4 | 4 | | 11 | 18 | 5 | 42 |
| 3/6 | 2 | 1 | 9 | 7 | 5 | 8 | 2 | 34 |
| 2/6 | | | | 3 | 2 | 5 | 1 | 11 |
| 1/6 | | 1 | | | 1 | | | 2 |
| 0 | | | 1 | 1 | 1 | | 1 | 4 |
| 总计 | 10 | 16 | 30 | 22 | 31 | 66 | 54 | 229 |

T＝保持对同事的信任是重要的

| | 0 | 1/6 | 2/6 | 3/6 | 4/6 | 5/6 | 1 | 总计 |
|---|---|---|---|---|---|---|---|---|
| 1 | 8 | 4 | 3 | 5 | 4 | 11 | 33 | 68 |
| 5/6 | 1 | 5 | 8 | 9 | 14 | 20 | 11 | 68 |
| 4/6 | | 2 | 5 | 9 | 5 | 16 | 5 | 42 |
| 3/6 | 1 | | | 7 | 8 | 11 | 7 | 34 |
| 2/6 | | | | 2 | 4 | 3 | 2 | 11 |
| 1/6 | | | 1 | | | 1 | | 2 |
| 0 | 1 | | | 1 | 1 | | 1 | 4 |
| 总计 | 11 | 11 | 17 | 33 | 36 | 62 | 59 | 229 |

I∨T

| | 0 | 1/6 | 2/6 | 3/6 | 4/6 | 5/6 | 1 | 总计 |
|---|---|---|---|---|---|---|---|---|
| 1 | | | | 1 | 1 | 13 | 53 | 68 |
| 5/6 | | | 1 | | 11 | 37 | 19 | 68 |
| 4/6 | | 1 | | | 10 | 24 | 7 | 42 |
| 3/6 | 1 | | | 4 | 7 | 14 | 8 | 34 |
| 2/6 | | | | 2 | 1 | 6 | 2 | 11 |
| 1/6 | | | 1 | | | 1 | | 2 |
| 0 | | | | 1 | 2 | | 1 | 4 |
| 总计 | 1 | 1 | 2 | 8 | 32 | 95 | 90 | 229 |

用模糊逻辑可能影响结果的四则运算法则，模型 C⇒I 的总真值达到了 229 中的 159.7，而模型 C⇒T 则达到了 162.3。平均

模型(即 C⇒[I+T]/2)也没有多大的改进,得到的真值总和为 166.7。然而 C⇒I∨T 的真值则高达 198.00,提高了 22%。一种多少更为常规一些的方法(尽管未必适当)会对照一个组合了 I 和 T 的预测线性【Linear】模型,用相关【Correlation】和回归【Regression】来评估 I∨T。I 和 C 之间的相关是0.114,而 T 和 C 之间的相关是 0.025。一个结合了两个预测因子的线性回归模型产生的多重相关 0.126,比零阶相关中的最大相关也没有多大改进。然而 I∨T 和 C 之间的相关却高达 0.430。

## 起源及发展

模糊集合理论是加利福尼亚大学伯克利分校的工程学教授卢特菲·扎德发明的,最初发表在他 1965 年的经典论文中。该理论的先声包括在 1920—1930 年代问世的多值逻辑和马克思·布莱克发表于 1937 年的开创性的用模糊概念推论的论文。模糊逻辑和可能性理论的基本观点产生于 1970 年代。在 1980 年代后期,模糊集合理论的公理化和模糊逻辑构建完成,而有关它们和概率框架的关系的争论也随之风生水起,至 20 世纪形成了明显的共识。部分是因为它的起源,模糊集合理论最早和最广泛的应用是在控制工程方面,在 1980 年代和 1990 年代,这些应用开始出现在其他有利可图的行业。在过去的 15 年,模糊集合理论和模糊逻辑融合了贝叶斯【Bayesian】概率,并和诸如在"软计算"方法中的神经网络【Neural Nets】和遗传算法这样的技术结合构建了智能系统。

在人文科学中的应用发展一直较慢。1970 年代,它首先出现在认知心理学中,与此同时,在社会科学中也有若干应用。在 1980 年代末,在从直觉研究到记忆人文地理研究的各种子领域内,出现了一些模糊集合理论的应用。不计相当多的报刊文章,至少出现了两本这方面的专著。尽管有关模糊集合理论在人文科学中的应用仍然见仁见智,但近年来有几位评论家认为,模糊集合理论的潜能至今仍未被人们所充分认识。

——Michael Smithson

(沈崇麟译校)

## 参考文献

Klir, G. J., & Folger, T. A. (1988). *Fuzzy sets, uncertainty, and information*. Englewood Cliffs, NJ: Prentice Hall.

Smithson, M. (1987). *Fuzzy set analysis for behavioral and social sciences*. New York: Springer-Verlag.

Zadeh, L. (1965). Fuzzy sets. *Information and Control*, *8*, 338-353.

Zetenyi, T. (Ed.). (1988). *Fuzzy sets in psychology*. Amsterdam: North-Holland.